

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:22:33 ; Search time 18.6885 seconds  
(without alignments)  
71.321 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEWPGSHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 469622

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	62	100.0	12	21	AA87461
2	62	100.0	15	10	AA87461
3	62	100.0	15	16	AA87461
4	62	100.0	21	21	AA87461
5	62	100.0	21	21	AA87461
6	62	100.0	23	16	AA87461
7	62	100.0	26	4	AA87461
8	62	100.0	41	6	AA87461
9	62	100.0	46	6	AA87461
10	62	100.0	47	4	AA87461
11	62	100.0	93	16	AA87461

12	62	100.0	93	20	AA87461
13	62	100.0	93	20	AA87461
14	62	100.0	93	21	AA87461
15	62	100.0	93	22	AA87461
16	47	75.8	15	21	AA87461
17	47	75.8	15	21	AA87461
18	46	74.2	15	22	AA87461
19	43	69.4	8	21	AA87461
20	39	62.9	7	21	AA87461
21	37	59.7	51	21	AA87461
22	35	56.5	57	22	AA87461
23	34	54.8	87	22	AA87461
24	34	54.8	92	21	AA87461
25	34	54.8	93	20	AA87461
26	33	53.2	49	22	AA87461
27	33	53.2	67	22	AA87461
28	33	53.2	71	20	AA87461
29	32	51.6	48	22	AA87461
30	32	51.6	57	22	AA87461
31	32	51.6	61	22	AA87461
32	32	51.6	62	22	AA87461
33	32	51.6	67	22	AA87461
34	32	51.6	78	22	AA87461
35	32	51.6	83	22	AA87461
36	32	51.6	88	22	AA87461
37	32	51.6	89	21	AA87461
38	32	51.6	90	22	AA87461
39	32	51.6	92	22	AA87461
40	32	51.6	95	22	AA87461
41	31	50.0	38	19	AA87461
42	31	50.0	38	21	AA87461
43	31	50.0	38	22	AA87461
44	31	50.0	44	22	AA87461
45	31	50.0	44	22	AA87461

#### ALIGNMENTS

RESULT 1  
ID AA87461 standard; peptide: 12 AA.  
AA87461:  
03-JUL-2000 (first entry)  
Cholera toxin B/enterotoxin B-derived peptide, SPO ID NO:3.  
Cholera toxin subunit B; CtxB, heat labile enterotoxin subunit B; EtxB;  
beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
adjuvant; immune disorder; diarrhoea.  
Vibrio cholerae.  
Escherichia coli.  
W0200014114-A1.  
16-MAR-2000.  
07-SEP-1999; 99NO-GB02970.  
07-SEP-1998; 98GB-0019484.  
(UYBR-) UNIV BRISTOL.  
Williams NA, Hirst TR.  
WPI: 2000-256943/22.  
Derivatives of Escherichia coli heat labile enterotoxins useful as  
immunomodulators and for treating diarrhoea and which do not bind the  
glycolipid receptor GM-1 -

XX Disclosure: Page 15; 62pp; English.  
 CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.

SO Sequence 12 AA;  
 Query Match 100.0%; Score 62; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12  
 |||||  
 DB 1 VEVPGSOHDSQ 12

RESULT 2  
 AAP93498  
 ID AAP93498 standard; protein; 15 AA.

AC AAP93498;

DT 03-MAY-1990 (first entry)

DE CTP3 epitope of the Cholera toxin B subunit.

KW CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;  
 KM vaccine; immunotherapy; ds;

PN WO8910967-A.

PD 16-NOV-1989.

PE 05-MAY-1989; 89WO-US01932.

PR 05-MAY-1988; 88US-0190570.

PA (PRAX-) PRAXIS BIOLOGICS INC.  
 PA (STRD ) LEYLAND STANDFORD JUNIOR UNIV.

PI Marjarian WR, Stocker BAD, Newton SMC;

DR WPI; 1989-356496/48.

DR N-PSDB; AAN92414.

XX New recombinant flagellin gene including sequence - for heterologous  
 PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.  
 PT of antibodies.

PS Disclosure: fig.4B; 137pp; English.

XX This sequence corresponds to the CTP3 epitope of the Cholera toxin B  
 CC subunit. The DNA sequence encoding this ligates to othersynthetic

CC oligonucleotides to form a new recombinant gene. This encodes  
 CC a flagellin fusion protein which can be used in vaccines for immuno-  
 CC therapy.

SO Sequence 15 AA;

Query Match 100.0%; Score 62; DB 10; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12  
 |||||  
 DB 1 VEVPGSOHDSQ 12

RESULT 3  
 AAR85125  
 ID AAR85125 standard; peptide; 15 AA.

AC AAR85125;

DT 13-JUN-1996 (first entry)

DE Cholera toxin B antigenic peptide fragment CTP3.

KW Conjugate; cholera; B toxin; peptide fragment; microparticulate;  
 KM inert carrier; modified silica; thyroglobulin; oral vaccine;  
 KM immunisation; infection; insoluble; digestive tract; antigen;  
 KM intestines; antibodies; secretory; Iga class.

OS Vibrio cholerae.

PN WO9529701-A1.

PD 09-NOV-1995.

PE 02-MAY-1995; 95WO-EP01661.

PR 03-MAY-1994; 94TL-0109519.

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Marks RS, Mirelman D, Sela M;

DR WPI; 1995-403805/51.

PT Vaccines for oral immunisation against infecting agents, e.g.

PT cholera - comprise a conjugate of an antigen of an infecting agent  
 PT covalently bound to micro:particulate inert carrier, e.g. modified  
 PT aldehyde silica

PS Claim 7; Page 25; 40pp; English.

XX A compsn. comprising a conjugate of an antigenic cholera B toxin  
 CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently  
 CC bound to a microparticulate inert carrier (e.g. modified silica or  
 CC thyroglobulin) can be used as an oral vaccine for immunisation  
 CC against cholera infection. The inert carrier is insoluble in the  
 CC digestive tract, allowing presentation of the antigen in the  
 CC intestines, where it will elicit antibodies mainly of the  
 CC secretory Iga class.

SO Sequence 15 AA;

Query Match 100.0%; Score 62; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12  
 |||||  
 DB 1 VEVPGSOHDSQ 12



```

RESULT 4
AAV87462
ID AAV87462 standard; peptide: 21 AA.
XX
AC AAV87462;
XX
DT 03-JUL-2000 (first entry)
XX
DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.
XX
KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;
KM beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW adjuvant; immune disorder; diarrhoea.
XX
OS Vibrio cholerae;
OS Escherichia coli.
XX
PN MO200014114-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-GB02970.
XX
PR 07-SEP-1998; 98GB-0019484.
XX
PA (UYBR-) UNIV BRISTOL.
XX
PI Williams NA, Hirst TR;
XX
DR WPI: 2000-256943/22.
XX
PT Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -
XX
XX
PS Disclosure: Page 15; 62pp; English.
XX
XX
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAV87460-Y87463
CC represent preferred peptides of the invention, AAV87460 being
CC particularly preferred.
XX
SQ Sequence 21 AA:
XX
Query Match 100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEPGSOHDSQ 12
DB 6 VEPGSOHDSQ 17
XX
RESULT 5
AAV87463
ID AAV87463 standard; peptide: 21 AA.

```

```

XX
AC AAV87463;
XX
DT 03-JUL-2000 (first entry)
XX
DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.
XX
KW Heat labile enterotoxin subunit B; EtxB;
KM beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW adjuvant; immune disorder; diarrhoea.
XX
OS Escherichia coli.
XX
PN MO200014114-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-GB02970.
XX
PR 07-SEP-1998; 98GB-0019484.
XX
PA (UYBR-) UNIV BRISTOL.
XX
PI Williams NA, Hirst TR;
XX
DR WPI: 2000-256943/22.
XX
PT Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -
XX
XX
PS Disclosure: Page 15; 62pp; English.
XX
XX
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAV87460-Y87463
CC represent preferred peptides of the invention, AAV87460 being
CC particularly preferred.
XX
SQ Sequence 21 AA:
XX
Query Match 100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEPGSOHDSQ 12
DB 6 VEPGSOHDSQ 17
XX
RESULT 6
AAV87448
ID AAV87448 standard; Protein: 23 AA.
XX
AC AAV87448;
XX
DT 18-MAR-1996 (first entry)

```

```

XX Residues 50-64 of cholera toxin B subunit and FimH 224-226.
DE FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; FimG; receptor binding site; PCR; amplify; ss.
XX Chimeric - Vibrio cholerae.
OS Chimeric - Escherichia coli.
FH Key Location/Qualifiers
FT Misc-difference 1
FT Peptide /note= "Represents FimH residue 224"
FT Peptide 2..4
FT Peptide /note= "Linker peptide"
FT Peptide 5..19
FT Peptide /note= "Cholera toxin B subunit 50-64"
FT Peptide 20..22
FT Misc-difference 23 /note= "Linker peptide"
FT Misc-difference /note= "Represents FimH residue 226"
XX WO9520657-A1.
XX 03-AUG-1995.
XX 27-JAN-1995; 95WO-DK00042.
XX 27-JAN-1994; 94US-0187166.
XX (GXRI-) GX BIOSYSTEMS AS.
XX Hasty DL, Klammm P, Molin S, Pallesen L, Sokurenko EV;
XX WPI; 1995-275442/36.
XX N-PSDB; AAQ93061.
XX Receptor specific bacterial adhesins - useful for targeting active
XX compounds and microbial cells to locations of receptors
XX Example 3; Page 58; 152pp; English.
XX This sequence is encoded by a fragment of the the plasmid pUPA93
XX which was used in the production of fimH fusion genes comprising
XX the cholera toxin B subunit inserted into the fimH gene. This insert
XX shows the inclusion of the B subunit into the fimH protein at position
XX 224-226. The chimeric genes were then opt. further modified by insertion
XX of the hepatitis B virus surface antigen pre-S2 region into a different
XX position of the fimH adhesin of type 1 fimbriae. Restriction site handles
XX (BglII-sites) were introduced into the fimH gene, and the foreign
XX epitopes are then inserted in-frame. In the selected positions the
XX function of the epitopes did not significantly alter the adhesive
XX on the surface of fimbriae on bacterial hosts illustrated the possibility
XX of using bacterial adhesins as general presenters of foreign antigens and
XX epitopes. These chimeric genes may be used in the production of variant
XX FimH adhesins which may be useful for targeting active compounds
XX and microbial cells to locations comprising selected receptors to which
XX the adhesins bind.
SQ Sequence 23 AA:
Query Match 100.0%; Score 62; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHIDSQ 12
DB 5 VEVPGSQHIDSQ 16

```

```

XX AAP30265;
AC 21-APR-1992 (first entry)
XX Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
DE carries an Arg at posns. 67 and 73.
XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
XX Vibrio cholerae.
OS EP95426-A.
XX 30-NOV-1983.
XX 26-MAY-1983; 83EP-0401052.
XX 26-MAY-1982; 82PR-0009167.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (INSP ) INST PASTEUR.
XX Milnaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
XX Guyongruaz A, Delmas A;
XX WPI; 1983-834645/49.
XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and
XX medicaments - effective against Escherichia coli and Vibrio
XX cholerae infections, are prepd. by solid phase peptide synthesis
XX Claim 7; Page 11; 13pp; French.
XX The inventors claim cholera toxin B1 subunit sequences which carry
XX Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
XX shown that Arg in posns. 35, 67 and 73 play an important role in fixing
XX these toxins to cell walls. The peptides are used in the treatment of,
XX and vaccination against, cholera infections and animal and human
XX infections due to E. coli (enterotoxin LT). The medicament may be
XX administered by oral, intraperitoneal, sub-cutaneous or intravenous
XX routes. For vaccines, pref. peptides having 15-30 AAs are injected
XX without carriers. Unit dose when used as a medicament is 50-500mg as
XX a vaccine 1-10mg of active cpd.
SQ Sequence 26 AA:
Query Match 100.0%; Score 62; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHIDSQ 12
DB 1 VEVPGSQHIDSQ 12

```

RESULT 7  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.

RESULT 8  
 AAP50439  
 ID AAP50439 standard; protein; 41 AA.

RESULT 7  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.

RESULT 7  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.

PD 20-JUN-1985.  
 XX 12-DEC-1984; 84MO-US02030.  
 XX 12-DEC-1983; 83US-0559469.  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 PA Houghten RA;  
 PI WPI; 1985-159230/26.  
 DR  
 XX  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 CC  
 XX  
 SQ Sequence 41 AA;  
 Query Match 100.0%; Score 62; DB 6; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 0.00018;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 DB 14 VEVPGSQHDSQ 25  
 RESULT 9  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX  
 AC AAP50436;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX  
 KM Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN W08502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PF 12-DEC-1984; 84MO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the

CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 CC  
 XX  
 SQ Sequence 46 AA;  
 Query Match 100.0%; Score 62; DB 6; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.00021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 DB 16 VEVPGSQHDSQ 27  
 RESULT 10  
 AAP30600  
 ID AAP30600 standard; Protein; 47 AA.  
 XX  
 AC AAP30600;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 35, 67 and 73.  
 XX  
 KM Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82FR-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCT.  
 PA (INSP ) INST PASTEUR.  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 DR WPI; 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 XX  
 PS Claim 8; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against, cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 CC  
 XX  
 SQ Sequence 47 AA;  
 Query Match 100.0%; Score 62; DB 4; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.00021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VEVPGSQHDSQ 12  
 DB 22 VEVPGSQHDSQ 33

```

RESULT 11
AA72545
ID AAR72545 standard; peptide: 93 AA.
XX
XX AAR72545;
XX
XX 28-NOV-1995 (first entry)
XX
XX ADP-ribosylating toxin (verotoxin-1 B-subunit).
DE
XX
XX ADP-ribosylating toxin; pertussis holotoxin; B-subunit;
XX active site; E. coli heat labile toxin; verotoxin-1;
XX Bordetella pertussis vaccines.
XX
XX Bacteria sp.
XX
XX EP646599-A.
XX
XX 05-APR-1995.
XX
XX 23-AUG-1994; 94EP-0306219.
XX
XX 24-AUG-1993; 93US-0110947.
XX 31-MAY-1994; 94US-0251121.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX (UYAL-) UNIV ALBERTA.
XX
XX Armstrong GD, Cockle SA, Hazes B, Klein MH, Loomore S;
XX Oomen R, Read RJ, Stein PE;
XX WPI; 1995-132623/18.
XX
XX New modified forms of pertussis holotoxin - developed using
XX crystalline forms of pertussis holotoxin and its complexes with
XX other molecules
XX
XX Disclosure; Fig 5; 54pp; English.
XX
XX AAR72540-R72545 are structurally equivalent B-subunits from three
XX ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat
XX labile toxin (LT), and verotoxin-1 (VT). The structural
XX information obtd. from these comparisons was used to identify
XX sites which contribute to PT's biological activity. By modifying
XX these sites the claimed PT mutants of the invention were produced,
XX they can be used in the development of vaccines against Bordetella
XX pertussis infection.
XX
XX Sequence 93 AA;
XX
XX Query Match 100.0%; Score 62; DB 16; Length 93;
XX Best Local Similarity 100.0%; Pred. No. 0.00045;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VEVPGSQHDSQ 12
XX |||||||||
XX 40 VEVPGSQHDSQ 51
XX
XX RESULT 12
XX AAY41816
XX ID AAY41816 standard; peptide: 93 AA.
XX
XX AAY41816;
XX
XX 08-DEC-1999 (first entry)
XX
XX Escherichia coli verotoxin-1 B-subunit.
XX
XX ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETR; DT;
XX three-dimensional structure; LT; immunoprotective; infection.
XX

```

```

XX
XX OS Escherichia coli.
XX
XX US5965385-A.
XX
XX 12-OCT-1999.
XX
XX 06-JUN-1995; 95US-0467974.
XX
XX 22-AUG-1994; 94US-0292968.
XX 24-AUG-1993; 93US-0110947.
XX 31-MAY-1994; 94US-0251121.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX (UYAL-) UNIV ALBERTA.
XX
XX Read RJ, Stein PE, Cockle SA, Loomore S, Klein MH, Armstrong GD;
XX Hazes B, Oomen RP;
XX WPI; 1999-579908/49.
XX
XX New method for producing modified pertussis holotoxin -
XX
XX Example 3; Fig 5; 41pp; English.
XX
XX A method has been developed of producing a modified pertussis holotoxin,
XX involving analysis of the 3-dimensional form of the crystalline
XX holotoxin. The pertussis holotoxin modification process comprises:
XX (1) identification of at least one amino acid (aa) residue of the
XX holotoxin for modification by analysing the 3-dimensional form of the
XX crystalline holotoxin, in relation to known information of the protein
XX structure and function; (2) effecting mutagenesis (by removing or
XX replacing a nucleotide sequence encoding at least one (aa) of a tox
XX operon; and (3) expressing mutant tox box in a Bordetella organism to
XX produce the modified holotoxin. This method is used for modifying
XX pertussis holotoxin, by studying its 3-dimensional crystalline
XX structure. Modifying the holotoxin, alters its biological properties.
XX By analysing the 3-dimensional crystalline structure of the pertussis
XX holotoxin, functional (aa) which affect biological properties of the
XX pertussis holotoxin can be identified. This can be used to predict (aa)
XX which contribute to the toxicity of the holotoxin to produce
XX immunoprotective, genetically-detoxified analogues of pertussis
XX holotoxin. The present sequence represents an ADP-ribosylating toxin
XX B-subunit peptide used in the exemplification of the present
XX invention.
XX
XX Sequence 93 AA;
XX
XX Query Match 100.0%; Score 62; DB 20; Length 93;
XX Best Local Similarity 100.0%; Pred. No. 0.00045;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 VEVPGSQHDSQ 12
XX |||||||||
XX 40 VEVPGSQHDSQ 51
XX
XX RESULT 13
XX AAW95226
XX ID AAW95226 standard; peptide: 93 AA.
XX
XX AAW95226;
XX
XX 16-MAR-1999 (first entry)
XX
XX E. coli heat-labile toxin (LT) beta-subunit sequence.
XX
XX Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;
XX enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;
XX structural analysis; interacting site; mitogenicity; adjuvanticity;
XX heat-labile; LT.
XX
XX Escherichia coli.
XX

```

```

XX  US5856122-A.
PN
XX
XX  05-JAN-1999.
PD
XX
XX  22-AUG-1994; 94US-0292968.
PF
XX
XX  22-AUG-1994; 94US-0292968.
PR
XX  24-AUG-1993; 93US-0110947.
PR
XX  31-MAY-1994; 94US-0251121.
XX
XX  (UYAL-) UNIV ALBERTA.
PA
XX
XX  Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI
XX  Oomen RP, Read RJ, Stein PE;
PI
XX  WPI: 1999-105104/09.
XX
XX  Modifications to e.g. enzymatic activity, mitogenicity and cell
PT  binding of pertussis holotoxin - by identifying interaction sites of
PT  a molecule with crystalline toxin and modifying the identified site
XX
XX  Example 3; Fig 5; 40pp; English.
PS
XX
XX  The invention relates to methods of preparing a pertussis holotoxin (PT)
CC  having a modified biological activity. One method comprises identifying
CC  at least 1 site in a PT that interacts with a molecule that is capable of
CC  forming a complex with the holotoxin and which molecule is an effector
CC  molecule which is an adenine nucleotide and which site contributes to
CC  toxicity, cell binding or enzymatic activity of PT. The functional
CC  interacting site(s) are identified by analysing the three dimensional
CC  structure of crystalline PT, determined by X-ray crystallography. The
CC  identified interacting site(s) are modified to alter toxicity, cell
CC  binding or enzyme activity of the PT. The methods can be used to alter a
CC  biological activity such as toxicity, enzymatic activity, mitogenicity,
CC  cell binding and adjuvant activity of the PT. The three-dimensional structure
CC  of PT have functional and/or structural resemblance to other bacterial
CC  toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
CC  heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
CC  sequence represents the beta-subunit of LT toxin.
CC
XX  Sequence 93 AA;
SQ
XX
XX  Query Match 100.0%; Score 62; DB 20; Length 93;
XX  Best Local Similarity 100.0%; Pred. NO. 0.00045;
XX  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 VEPGSHIDSO 12
XX  |||||
XX  DB 40 VEPGSHIDSO 51
XX
XX  RESULT 14
XX  AAY68365
XX  ID AAY68365 standard; Peptide; 93 AA.
XX
XX  AC AAY68365;
XX
XX  DT 17-APR-2000 (first entry)
XX
XX  DE Heat labile toxin B subunit SEQ ID NO:26.
XX
XX  KW Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;
XX  diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;
XX  infection; crystal structure; X-ray crystallography; detoxification;
XX  immunogenic.
XX
XX  OS Escherichia coli.
XX
XX  PN US6018022-A.
XX
XX  PD 25-JAN-2000.
XX

```

```

PF  06-JUN-1995; 95US-0467976.
XX
XX  22-AUG-1994; 94US-0292968.
PR
XX  24-AUG-1993; 93US-0110947.
PR
XX  31-MAY-1994; 94US-0251121.
XX
XX  (CONN-) CONNAUGHT LAB LTD.
XX  (UYAL-) UNIV ALBERTA.
PA
XX
XX  Read RJ, Oomen RP, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
PI
XX  Hazes B, Stein PE;
PI
XX  WPI: 2000-136703/12.
XX
XX  Crystalline form of isolated pertussis holotoxin useful in studying
PT  proteins which have functional resemblance -
XX
XX  Example 3; Fig 5; 42pp; English.
PS
XX
XX  The present invention describes a crystalline form of isolated
CC  pertussis holotoxin, in which the molecules of pertussis toxin have
CC  a three dimensional structure represented in the specification,
CC  complexed with a polysaccharide molecule capable of forming a complex
CC  with the holotoxin. The crystalline form of the pertussis holotoxin
CC  can be used in a comparison with other proteins which have functional
CC  resemblance to pertussis holotoxin with the aim of modifying other
CC  proteins. Identifying the unknown sites of toxicity by comparison
CC  with the three dimensional structure of pertussis holotoxin provides a
CC  technique for detoxification of toxins to produce useful immunogenic
CC  but non-toxic analogues. It can also be used as a primary standard to
CC  measure the quantity, purity or efficacy of less pure compositions
CC  containing pertussis toxin. AAY68340 to AAY68385 represent peptides
CC  used in the exemplification of the present invention.
CC
XX  Sequence 93 AA;
SQ
XX
XX  Query Match 100.0%; Score 62; DB 21; Length 93;
XX  Best Local Similarity 100.0%; Pred. NO. 0.00045;
XX  Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 VEPGSHIDSO 12
XX  |||||
XX  DB 40 VEPGSHIDSO 51
XX
XX  RESULT 15
XX  AAB66239
XX  ID AAB66239 standard; Protein; 93 AA.
XX
XX  AC AAB66239;
XX
XX  DT 03-APR-2001 (first entry)
XX
XX  DE E coli verotoxin-1 B subunit SEQ ID NO: 26.
XX
XX  KW Pertussis toxin; crystal structure; whooping cough; biological activity;
XX  lymphocytosis-promoting factor; histamine-sensitising factor;
XX  islet-activating protein.
XX
XX  OS Escherichia coli.
XX
XX  PN US6168928-B1.
XX
XX  PD 02-JAN-2001.
XX
XX  PF 21-MAY-1998; 98US-0082514.
XX
XX  PR 22-AUG-1994; 94US-0292968.
XX  24-AUG-1993; 93US-0110947.
XX  31-MAY-1994; 94US-0251121.
XX
XX  (CONN-) CONNAUGHT LAB LTD.
XX

```

PI Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Stein PE;  
 XX  
 DR WPI; 2001-122260/13.

XX  
 PT Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site -  
 XX

PS Example 3; Fig 5; 41pp; English.

XX  
 CC The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitizing  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.

CC  
 SQ Sequence 93 AA;

Query Match 100.0%; Score 62; DB 22; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.00045;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSGQHDSQ 12

|||||

DB 40 VEPGSGQHDSQ 51

Search completed: October 24, 2002, 15:35:02  
 Job time : 19.6885 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:33:23 : Search time 6.88525 Seconds  
(without alignments)  
42.570 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEVPGSQHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 177902

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCrus.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	93	2	US-08-292-968-26
2	62	100.0	93	2	US-08-467-974-26
3	62	100.0	93	2	US-08-467-974-26
4	62	100.0	93	3	US-08-467-976-26
5	62	100.0	93	4	US-09-082-514-26
6	34	54.8	93	4	US-09-069-023-15
7	31	50.0	38	2	US-08-726-306A-101
8	31	50.0	41	6	5208144-18
9	30.5	49.2	15	1	US-08-477-108A-11
10	30	48.4	11	2	US-08-747-410-12
11	30	48.4	57	5	PCT-US92-07813-26
12	30	48.4	98	3	US-09-003-708A-6
13	30	48.4	98	4	US-09-383-212-6
14	29	46.8	20	2	US-08-934-915-14
15	29	46.8	34	4	US-09-314-268-125
16	28.5	46.0	75	1	US-08-239-256-3
17	28.5	46.0	76	6	5177193-5
18	28	45.2	18	3	US-08-817-926-10
19	28	45.2	18	3	US-08-926-842B-56
20	28	45.2	39	4	US-09-450-072-78
21	28	45.2	99	4	US-08-936-165A-475
22	28	45.2	100	4	US-09-031-962D-20
23	27	43.5	13	4	US-08-540-922D-5
24	27	43.5	14	1	US-08-258-851-7
25	27	43.5	16	1	US-07-789-912-5
26	27	43.5	48	1	US-07-763-860-6
27	27	43.5	49	4	US-08-975-080-24

28	27	43.5	52	1	US-07-689-693B-24	Sequence 24, Appl
29	27	43.5	56	1	US-07-763-860-7	Sequence 7, Appl
30	27	43.5	57	1	US-07-763-860-8	Sequence 8, Appl
31	27	43.5	60	1	US-07-763-860-9	Sequence 9, Appl
32	27	43.5	62	1	US-07-763-860-10	Sequence 10, Appl
33	27	43.5	64	1	US-07-763-860-11	Sequence 11, Appl
34	27	43.5	64	1	US-07-763-860-13	Sequence 13, Appl
35	27	43.5	65	1	US-07-763-860-12	Sequence 12, Appl
36	27	43.5	66	1	US-07-763-860-5	Sequence 5, Appl
37	27	43.5	68	1	US-07-689-693B-11	Sequence 11, Appl
38	27	43.5	68	2	US-08-511-485-29	Sequence 29, Appl
39	27	43.5	90	4	US-09-227-357-635	Sequence 635, App
40	26.5	42.7	35	1	US-08-418-893D-16	Sequence 16, Appl
41	26	41.9	8	2	US-08-318-856A-30	Sequence 30, Appl
42	26	41.9	9	2	US-08-318-856A-22	Sequence 22, Appl
43	26	41.9	9	4	US-09-139-802-193	Sequence 193, App
44	26	41.9	9	4	US-09-095-443-8	Sequence 8, Appl
45	26	41.9	10	2	US-08-318-856A-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 5856122

GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/292,968  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO.: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
|||||  
DB 40 VEVPGSQHDSQ 51

RESULT 2  
US-08-467-974-26

; Sequence 26, Application US/08467974

; Patent No. 5965385

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,974

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/467,536

; FILING DATE: 06-JUN-1995

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-454 MIS.vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1153

; TELEFAX: (416) 595-1153

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-974-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12

DB 40 VEVPGSQHDSQ 51

RESULT 3  
US-08-467-536-26

; Sequence 26, Application US/08467536

; Patent No. 5977304

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,536

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-455 MIS.vg

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1153

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12

DB 40 VEVPGSQHDSQ 51

RESULT 4  
US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022  
; GENERAL INFORMATION:



```

: APPLICANT: READ, Randy J.
: APPLICANT: STEIN, Penelope E.
: APPLICANT: COCKLE, Stephen A.
: APPLICANT: COOMEN, Raymond P.
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: KLEIN, Michel H.
: APPLICANT: ARMSTRONG, Glen D.
: APPLICANT: HAZES, Bart
: TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467,976
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/292,968
: FILING DATE: 22-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/251,121
: FILING DATE: 31-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/110,947
: FILING DATE: 24-AUG-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-453 MTS:vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-467-976-26

Query Match          100.0%; Score 62; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEPGSHIDSQ 12
      |||||||
DB      40 VEPGSHIDSQ 51

RESULT 5
US-09-082-514-26
: Sequence 26, Application US/09082514
: Patent No. 6168928
: GENERAL INFORMATION:
: APPLICANT: READ, Randy J.
: APPLICANT: STEIN, Penelope E.
: APPLICANT: COCKLE, Stephen A.
: APPLICANT: COOMEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: APPLICANT: ARMSTRONG, Glen D.
: APPLICANT: HAZES, Bart
: TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
: NUMBER OF SEQUENCES: 46
```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,514
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/292,968
: FILING DATE: 24-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-810
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-082-514-26

Query Match          100.0%; Score 62; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEPGSHIDSQ 12
      |||||||
DB      40 VEPGSHIDSQ 51

RESULT 6
US-09-069-023-15
: Sequence 15, Application US/09069023A
: Patent No. 6348573
: GENERAL INFORMATION:
: APPLICANT: Nunez, Gabriel
: APPLICANT: Inohara, Naohiro
: APPLICANT: KOSAKI, Takeyoshi
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
: FILE REFERENCE: UM-03333
: CURRENT APPLICATION NUMBER: US/09/069,023A
: CURRENT FILING DATE: 1998-04-27
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 15
: LENGTH: 93
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-069-023-15

Query Match          54.8%; Score 34; DB 4; Length 93;
Best Local Similarity 71.4%; Pred. No. 32;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      3 VPGSHI 9
      :|||||
DB      2 MPGSHV 8
```

RESULT 7  
US-08-726-306A-101  
; Sequence 101, Application US/08726306A  
; Patent No. 5958684  
; GENERAL INFORMATION:  
; APPLICANT: van Leeuwen, Frederik Willem  
; APPLICANT: Barbach, Johannes Peter Henri  
; APPLICANT: Grosveld, Franklin G.  
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS  
; NUMBER OF SEQUENCES: 189  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1 Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/726,306A  
; FILING DATE: 02-Oct-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 95/20080.4  
; FILING DATE: 02-Oct-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/009,832  
; FILING DATE: 01-Jan-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 101:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-726-306A-101  
Query Match 50.0%; Score 31; DB 2; Length 38;  
Best Local Similarity 60.0%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 EYVGSQHIDS 11  
DB 15 EGPGRHDS 24  
RESULT 8  
5208144-18  
; Patent No. 5208144  
; APPLICANT: SMITH, JOHN A.; RAYCHOMDHURY, RAKTIMA; NILES, JOHN L.  
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA  
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 42  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/396,697  
; FILING DATE: 22-AUG-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 313,682  
; FILING DATE: 22-FEB-1989  
; APPLICATION NUMBER: 235,211  
; FILING DATE: 23-AUG-1988  
; SEQ ID NO: 18:  
; LENGTH: 41

5208144-18  
Query Match 50.0%; Score 31; DB 6; Length 41;  
Best Local Similarity 50.0%; Pred. No. 44;  
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 3 VEGSQHDSQ 12  
DB 23 IPEGWHCDRQ 32  
RESULT 9  
US-08-477-108A-11  
; Sequence 11, Application US/08477108A  
; Patent No. 5801001  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; APPLICANT: Zou, Zhiqiang  
; APPLICANT: Antosiewicz, Anthony  
; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; COMPUTER: IBM PS/2 Model 502 or 555X  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,108A  
; FILING DATE: June 7, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/121,714  
; FILING DATE: 09/01/93  
; APPLICATION NUMBER: 07/938,823  
; FILING DATE: 09/01/92  
; APPLICATION NUMBER: 07/844,296  
; FILING DATE: 02/28/92  
; APPLICATION NUMBER: 07/662,216  
; FILING DATE: 02/28/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 06570/002002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-477-108A-11  
Query Match 49.2%; Score 30.5; DB 1; Length 15;  
Best Local Similarity 53.8%; Pred. No. 18;  
Matches 7; Conservative 2; Mismatches 1; Indels 3; Gaps 1;  
QY 1 VEVPGS---QHID 10  
DB 1 IEVPGARIHQHD 13  
RESULT 10

US-08-747-410-12  
: Sequence 12, Application US/08747410  
: Patent No. 5993820  
: GENERAL INFORMATION:  
: APPLICANT: BAGDASARIAN, Michael  
: APPLICANT: IRELAND, James  
: TITLE OF INVENTION: CHIMERIC I/FB VACCINES  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
: STREET: 3100 No. 5993820west Center, 90 South Seventh St  
: CITY: Minneapolis  
: STATE: MN  
: COUNTRY: USA  
: ZIP: 55402  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: DOS  
: SOFTWARE: FastSeq Version 1.5  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/747,410  
: FILING DATE: 12-NOV-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Kettelberger, Denise M  
: REGISTRATION NUMBER: 33,924  
: REFERENCE/DOCKET NUMBER: 11526.1-US-01  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 612/371-5268  
: TELEFAX: 612/332-9081  
: TELEX:  
: INFORMATION FOR SEQ ID NO: 12:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 11 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: HYPOTHETICAL: NO  
: ANTI-SENSE: NO  
: FRAGMENT TYPE: internal  
: ORIGINAL SOURCE:  
: US-08-747-410-12

Query Match 48.4%; Score 30; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSG 6  
| | | | |  
DB 1 VEPGSG 6

RESULT 11  
PCT-US92-07813-26  
: Sequence 26, Application PC/US9207813  
: GENERAL INFORMATION:  
: APPLICANT: LESNIEWSKI, RICHARD R.  
: APPLICANT: LEUNG, TAT K.  
: TITLE OF INVENTION: HEPATITIS C ASSAY  
: NUMBER OF SEQUENCES: 27  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: ABBOTT LABORATORIES CHAD377/AP6D  
: STREET: ONE ABBOTT PARK ROAD  
: CITY: ABBOTT PARK  
: STATE: ILLINOIS  
: COUNTRY: U.S.A.  
: ZIP: 60064-3500  
: COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: PCT/US92/07813  
: FILING DATE: 19920916  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: FOREMSKIP, PRISCILLA E.  
: REGISTRATION NUMBER: 33,207  
: REFERENCE/DOCKET NUMBER: 4767.P3.03  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 708-937-6365  
: TELEFAX: 708-937-9556  
: INFORMATION FOR SEQ ID NO: 26:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 57 amino acids  
: TYPE: AMINO ACID  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: PCT-US92-07813-26

Query Match 48.4%; Score 30; DB 5; Length 57;  
Best Local Similarity 60.0%; Pred. No. 96;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEPGSGHID 10  
| | | | |  
DB 35 VMTNGSQHIN 44

RESULT 12  
US-09-003-708A-6  
: Sequence 6, Application US/09003708A  
: Patent No. 6010882  
: GENERAL INFORMATION:  
: APPLICANT: WU, Jen-Leih  
: APPLICANT: CHEN, Jyh-Yih  
: APPLICANT: CHANG, Chi-Yao  
: TITLE OF INVENTION: Production of Biologically  
: TITLE OF INVENTION: Active Recombinant Insulin-Like Growth Factor II  
: NUMBER OF SEQUENCES: 6  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Snider & Chao, L.L.P.  
: STREET: P.O. Box 27613  
: CITY: Washington, D.C.  
: STATE: DC  
: COUNTRY: USA  
: ZIP: 20038  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
: MEDIUM TYPE: storage  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: MS-DOS  
: SOFTWARE: WordPerfect 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/003,708A  
: FILING DATE: 01/07/98  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 60/034/736  
: FILING DATE: January 10, 1997  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Ronald R. Snider  
: REGISTRATION NUMBER: 24,962  
: REFERENCE/DOCKET NUMBER: ACA-001  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202)-347-2600  
: TELEFAX: (202)-347-7870  
: INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: Unknown  
MOLECULE TYPE: polypeptide  
HYPOTHETICAL: NO  
US-09-003-708A-6

Query Match 48.4%; Score 30; DB 3; Length 98;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVGSQHI 9  
DB 20 EYVKKQHV 27

RESULT 13  
US-09-383-212-6  
Sequence 6, Application US/09383212  
Patent No. 6235874

GENERAL INFORMATION:  
APPLICANT: WU, Jen-Leih  
APPLICANT: CHEN, Jyh-Yih  
TITLE OF INVENTION: Production of Biologically Active  
TITLE OF INVENTION: Recombinant Insulin-Like Growth Factor II  
TITLE OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Snider & Chao, L.L.P.  
STREET: P.O. Box 27613  
CITY: Washington, D.C.  
STATE: DC  
COUNTRY: USA  
ZIP: 20038  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
MEDIUM TYPE: Storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/383,212  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/003,708  
FILING DATE: 01/07/98

APPLICATION NUMBER: 60/034/736  
FILING DATE: January 10, 1997

APPLICATION NUMBER: 09/003/708  
FILING DATE: January 7, 1998

ATTORNEY/AGENT INFORMATION:  
NAME: Pei-Fei Chao  
REGISTRATION NUMBER: 43,538

REFERENCE/DOCKET NUMBER: ACA-001-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)-347-2600

TELEFAX: (202)-347-7870  
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid

TOPOLOGY: Unknown  
MOLECULE TYPE: polypeptide  
HYPOTHETICAL: NO

US-09-383-212-6

Query Match 48.4%; Score 30; DB 4; Length 98;  
Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYVGSQHI 9

DB 20 EYVKKQHV 27

RESULT 14  
US-08-934-915-14  
Sequence 14, Application US/08934915  
Patent No. 5932412

GENERAL INFORMATION:  
APPLICANT: DILLNER, JOAKIM  
APPLICANT: DILLNER, LENA

APPLICANT: CHENG, HWEE-MING  
TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN

TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,  
TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,  
TITLE OF INVENTION: USEFUL IN IMMUNOSSAY FOR

TITLE OF INVENTION: DIAGNOSTIC PURPOSES  
NUMBER OF SEQUENCES: 193  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: MASON & ASSOCIATES, P.A.  
STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500  
CITY: CLEARWATER

STATE: FLORIDA  
COUNTRY: U.S.A.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 3.0

SOFTWARE: Microsoft Word 6.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,915  
FILING DATE: 22-SEP-1997  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/949,836  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: LOUISE A. Foulch  
REGISTRATION NUMBER: 37,133

REFERENCE/DOCKET NUMBER: 1946.6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 813-538-3800

TELEFAX: 813-538-3820  
TELEX:

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-934-915-14

Query Match 46.8%; Score 29; DB 2; Length 20;  
Best Local Similarity 66.7%; Pred. No. 45;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 GSGQHD 10  
DB 5 GSEHVD 10

RESULT 15  
US-09-314-268-125  
Sequence 125, Application US/09314268  
Patent No. 6346377

GENERAL INFORMATION:  
APPLICANT: DOORBAT, John  
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

FILE REFERENCE: 3789/80902  
CURRENT APPLICATION NUMBER: US/09/314,268  
CURRENT FILING DATE: 1999-03-19

EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18  
 ; NUMBER OF SEQ ID NOS: 179  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 125  
 ; LENGTH: 34  
 ; TYPE: PRT  
 ; ORGANISM: Human papillomavirus type 11  
 US-09-314-268-125

Query Match 46.88; Score 29; DB 4; Length 34;  
 Best Local Similarity 66.78; Pred. NO. 81;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 GSOHD 10  
 11:1:1  
 Db 17 GSEHD 22

Search completed: October 24, 2002, 15:38:22  
 Job time : 7.88525 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:09 : Search time 79.8689 Seconds  
(without alignments)  
52.884 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62  
Sequence: 1 VEPGSGHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 2671256

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCTUS.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US05.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US080.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US081.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US082.COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US60.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	12	21	US-09-786-648-3
2	62	100.0	15	11	US-08-732-371-1
3	62	100.0	15	11	US-08-732-371A-1
4	62	100.0	21	21	US-09-786-648-4
5	62	100.0	21	21	US-09-786-648-5
6	62	100.0	93	5	US-08-110-947-10
7	62	100.0	93	5	US-08-110-947A-26

8	62	100.0	93	6	US-08-251-121-26	Sequence 26, Appl
9	39	62.9	7	21	US-09-786-648-2	Sequence 2, Appl1
10	39	62.9	91	19	US-09-595-298A-470	Sequence 470, App
11	36	58.1	51	22	US-09-866-066-35	Sequence 35, Appl
12	35	56.5	75	18	US-09-465-854-18	Sequence 18, Appl1
13	35	56.5	75	26	US-60-112-736-8	Sequence 8, Appl1
14	35	56.5	91	12	US-08-827-356-4889	Sequence 4889, Ap
15	35	56.5	91	20	US-09-611-529-6316	Sequence 6316, Ap
16	35	56.5	92	26	US-60-143-992-917	Sequence 917, App
17	35	56.5	92	26	US-60-145-138-477	Sequence 477, App
18	35	56.5	94	1	PCT-US01-14827-11193	Sequence 11193, A
19	34	54.8	41	12	US-08-888-755-16	Sequence 16, Appl1
20	34	54.8	41	12	US-08-888-755A-16	Sequence 16, Appl1
21	34	54.8	41	20	US-09-688-051-2967	Sequence 2967, Ap
22	34	54.8	46	24	US-10-029-386-29640	Sequence 29640, A
23	34	54.8	84	20	US-09-617-682A-5651	Sequence 5651, Ap
24	34	54.8	87	1	PCT-US01-02723-224	Sequence 224, App
25	34	54.8	87	21	US-09-707-351-224	Sequence 352, App
26	34	54.8	93	13	US-08-905-144-352	Sequence 583, App
27	34	54.8	96	21	US-09-760-495-583	Sequence 6, Appl1
28	33	53.2	49	18	US-09-455-895-6	Sequence 6, Appl1
29	33	53.2	49	21	US-09-728-912-6	Sequence 2501, Ap
30	33	53.2	67	1	PCT-US01-03800A-2501	Sequence 389, App
31	33	53.2	71	17	US-09-397-945-448	Sequence 448, App
32	33	53.2	82	26	US-60-186-656-1008	Sequence 1008, Ap
33	33	53.2	84	26	US-60-188-162-4087	Sequence 4087, Ap
34	33	53.2	86	20	US-09-620-111B-1049	Sequence 1049, Ap
35	33	53.2	89	26	US-60-182-898-229	Sequence 229, App
36	33	53.2	93	26	US-60-182-898-268	Sequence 268, App
37	33	53.2	96	23	US-09-902-540-15400	Sequence 15400, A
38	33	53.2	97	21	US-09-758-440-645	Sequence 645, App
39	33	53.2	97	26	US-60-200-109-1132	Sequence 1132, Ap
40	33	53.2	98	26	US-60-200-109-1129	Sequence 1129, Ap
41	33	53.2	98	26	PCT-US00-24802-3025	Sequence 3025, Ap
42	32	51.6	15	1	PCT-US00-24802-3183	Sequence 3183, Ap
43	32	51.6	15	16	US-09-239-043-2176	Sequence 2176, Ap
44	32	51.6	15	16	US-09-239-043-2176	Sequence 2225, Ap
45	32	51.6	15	16	US-09-239-043-2225	

#### ALIGNMENTS

RESULT 1

US-09-786-648-3

Sequence 3, Application US/09786648

GENERAL INFORMATION:

APPLICANT: Williams, Neil Andrew

TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci

FILE REFERENCE: 7438

CURRENT APPLICATION NUMBER: US/09/786,648

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: PCT/GB99/02970

PRIOR FILING DATE: 1999-09-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: MS DOS

SEQ ID NO 3

LENGTH: 12

TYPE: PRT

ORGANISM: E. coli

FEATURE:

LOCATION: 50...61

OTHER INFORMATION: isolated or synthetic EtxB beta4-alpha2 loop fragment derivabl

US-09-786-648-3

Query Match 100.0%; Score 62; DB 21; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|||||  
Db 1 VEVPSQHDSQ 12

## RESULT 2

US-08-732-371-1  
; Sequence 1, Application US/08732371

; GENERAL INFORMATION:

; APPLICANT: MIRELMAN, David

; APPLICANT: MARKS, Robert S.

; APPLICANT: SELA, Michael

; TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST

; TITLE OF INVENTION: INFECTING AGENTS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/732,371

; FILING DATE: 09-JAN-1997

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: IL 109519

; FILING DATE: 03-MAY-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: YUN, Allen C.

; REGISTRATION NUMBER: 37,971

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-732-371-1

Query Match 100.0%; Score 62; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||

Db 1 VEVPSQHDSQ 12

## RESULT 3

US-08-732-371A-1  
; Sequence 1, Application US/08732371A

; GENERAL INFORMATION:

; APPLICANT: MIRELMAN, David

; APPLICANT: MARKS, Robert S.

; APPLICANT: SELA, Michael

; TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST

; TITLE OF INVENTION: INFECTING AGENTS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

ZIP: 20004  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/732,371A

; FILING DATE: 09-JAN-1997

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: IL 109519

; FILING DATE: 03-MAY-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: YUN, Allen C.

; REGISTRATION NUMBER: 37,971

; REFERENCE/DOCKET NUMBER: MIRELMAN-3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-732-371A-1

Query Match 100.0%; Score 62; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||

Db 1 VEVPSQHDSQ 12

## RESULT 4

US-09-786-648-4  
; Sequence 4, Application US/09786648

; GENERAL INFORMATION:

; APPLICANT: Williams, Neil Andrew

; APPLICANT: Hirst, Timothy Raymond

; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacc

; TITLE OF INVENTION: Adjuvants

; FILE REFERENCE: 7438

; CURRENT APPLICATION NUMBER: US/09/786,648

; PRIOR FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: PCT/GB99/02970

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: MS DOS

; SEQ ID NO 4

; LENGTH: 21

; TYPE: PPT

; ORGANISM: E. coli

; FEATURE:

; LOCATION: 45...65

; OTHER INFORMATION: Isolated or synthetic Etxb beta4-alpha2 loop fragment derivabl

; OTHER INFORMATION: human variant E. coli

; US-09-786-648-4

Query Match 100.0%; Score 62; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.00079;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||

Db 6 VEVPSQHDSQ 17

## RESULT 5

US-09-786-648-5



```

; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic EtXB beta4-alpha2 loop fragment derivable f
; OTHER INFORMATION: porcine E. coli
US-09-786-648-5

Query Match          100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VEPGSOHDSQ 12
Db      6 VEPGSOHDSQ 17

RESULT 6
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Hwy.
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
US-08-110-947-10

Query Match          100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VEPGSOHDSQ 12
Db      40 VEPGSOHDSQ 51

RESULT 7
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947A-26

Query Match          100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VEPGSOHDSQ 12
Db      40 VEPGSOHDSQ 51

RESULT 8
US-08-251-121-26
; Sequence 26, Application US/08251121
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: COOMEN, Raymond P.

```

APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/251,121  
FILING DATE: 31-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-335  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-251-121-26

Query Match 100.0%; Score 62; DB 6; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.0044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
DB 40 VEPGSOHDSQ 51

RESULT 9  
US-09-786-648-2  
; Sequence 2, Application US/097866648  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Neil Andrew  
; APPLICANT: Hirst, Timothy Raymond  
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine  
; TITLE OF INVENTION: Adjuvants  
; FILE REFERENCE: 7438  
; CURRENT APPLICATION NUMBER: US/09/786,648  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: PCT/G899/02970  
; PRIOR FILING DATE: 1999-09-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: MS DOS  
; SEQ ID NO 2  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: E. coli  
; FEATURE:  
; LOCATION: 51..57  
; OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivable  
; OTHER INFORMATION: human variant E. coli  
US-09-786-648-2

Query Match 62.9%; Score 39; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3.2e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVPGSOH 8  
DB 1 EVPGSOH 7

RESULT 10  
US-09-595-298A-470  
; Sequence 470, Application US/09595298A  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai  
; APPLICANT: BROVER, Vyacheslav  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 2750-0953P  
; CURRENT APPLICATION NUMBER: US/09/595,298A  
; CURRENT FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 2756  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 470  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Arabidopsis Thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: (1)..(91)  
; OTHER INFORMATION: Ceres Seq. ID no. 1024399  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: Xaa is any aa, unknown or other  
US-09-595-298A-470

Query Match 62.9%; Score 39; DB 19; Length 91;  
Best Local Similarity 50.0%; Pred. No. 65;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
DB 41 VKVPGKOHSEK 52

RESULT 11  
US-09-866-066-35  
; Sequence 35, Application US/09866066  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Christopher  
; APPLICANT: Roberts, Steve  
; APPLICANT: Ruble, Cara  
; APPLICANT: Gotow, Lisa  
; APPLICANT: Karnovsky, Alla  
; TITLE OF INVENTION: Human Ion Channels  
; FILE REFERENCE: 00229 US1  
; CURRENT APPLICATION NUMBER: US/09/866,066  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/207,152  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/207,257  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/207,119  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 35  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-866-066-35

Query Match 58.1%; Score 36; DB 22; Length 51;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYPSGSHD 10  
|:|||||  
Db 8 EIPGSSHMD 16

## RESULT 12

US-09-465-854-18  
; Sequence 18, Application US/09465854  
; GENERAL INFORMATION:  
; APPLICANT: Kinney, Tony  
; APPLICANT: Lane, Tammy M.  
; APPLICANT: Rafalski, Antoni  
; TITLE OF INVENTION: Plant Purine Biosynthetic Genes  
; FILE REFERENCE: BB1323 US NA  
; CURRENT APPLICATION NUMBER: US/09/465,854  
; CURRENT FILING DATE: 1999-12-17  
; EARLIER APPLICATION NUMBER: 60/112,736  
; EARLIER FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-09-465-854-18

Query Match 56.5%; Score 35; DB 18; Length 75;  
Best Local Similarity 75.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYPSGSHD 9  
:|||||  
Db 5 DVPGSNHI 12

## RESULT 13

US-60-112-736-8  
; Sequence 8, Application US/60112736  
; GENERAL INFORMATION:  
; APPLICANT: Kinney, Anthony J.  
; APPLICANT: Lane, Tammy M.  
; APPLICANT: Rafalski, J. Antoni  
; TITLE OF INVENTION: Plant Purine Biosynthetic Genes  
; FILE REFERENCE: BB-1323-P1  
; CURRENT APPLICATION NUMBER: US/60/112,736  
; CURRENT FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Microsoft Word 7.0A  
; SEQ ID NO 8  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Triticum aestivum  
US-60-112-736-8

Query Match 56.5%; Score 35; DB 26; Length 75;  
Best Local Similarity 75.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYPSGSHD 9  
:|||||  
Db 5 DVPGSNHI 12

## RESULT 14

US-08-827-356-4889  
; Sequence 4889, Application US/08827356  
; GENERAL INFORMATION:  
; APPLICANT: George H. Shimer, Jr.  
; APPLICANT: George H. Miller  
; APPLICANT: Roberta S. Hare  
; APPLICANT: Karen J. Shaw

TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 5574  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schering-Plough Corporation  
; STREET: 2000 Galloping Hill Road  
; CITY: Kenilworth  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07033-0530

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentlin

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,356  
; FILING DATE: 01-APR-1997

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/014,477  
; FILING DATE: 01-APR-1996

APPLICATION NUMBER: 60/016,743  
; FILING DATE: 02-MAY-1996  
; APPLICATION NUMBER: 60/020,016  
; FILING DATE: 14-JUN-1996

INFORMATION FOR SEQ ID NO: 4889:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 91 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: YES

ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1...91

US-08-827-356-4889

Query Match 56.5%; Score 35; DB 12; Length 91;  
Best Local Similarity 36.4%; Pred. No. 3.5e+02;  
Matches 4; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYPSGSHDSD 12  
:|||||  
Db 17 QIPGPNHVESR 27

## RESULT 15

US-09-611-529-6316  
; Sequence 6316, Application US/09611529  
; GENERAL INFORMATION:  
; APPLICANT: George H. Shimer, Jr.  
; APPLICANT: George H. Miller  
; APPLICANT: Roberta S. Hare  
; APPLICANT: Karen J. Shaw

TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods  
; FILE REFERENCE: 1034/1C963051  
; CURRENT APPLICATION NUMBER: US/09/611,529  
; CURRENT FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/417,811  
; PRIOR FILING DATE: 1999-10-14

PRIOR APPLICATION NUMBER: US 09/353,718  
; PRIOR FILING DATE: 1999-07-14

PRIOR APPLICATION NUMBER: US 09/266,557  
; PRIOR FILING DATE: 1999-03-11

PRIOR APPLICATION NUMBER: US 09/266,556  
; PRIOR FILING DATE: 1999-03-11

PRIOR APPLICATION NUMBER: US 09/266,555  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,542  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,541

```
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,334
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,221
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,137
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,082
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,081
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,079
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,913
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/035,744
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 08/827,356
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 08/831,156
; PRIOR FILING DATE: 1997-04-01
; PRIOR APPLICATION NUMBER: US 60/014,477
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: US 60/016,743
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: US 60/020,016
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 6316
; LENGTH: 91
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-611-529-6316
```

```
Query Match 56.5%; Score 35; DB 20; Length 91;
Best local Similarity 36.4%; Pred. No. 3.5e+02;
Matches 4; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 2 EVPGSQHDSQ 12
: : | | | : : | :
Db 17 QIPGPNHVESR 27
```

Search completed: October 24, 2002, 15:45:15  
Job time : 80.8689 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:39 : Search time 26.1639 Seconds  
(without alignments)  
137.008 Million cell updates/sec

Title: US-09-786-648-3

Sequence: 1 VEPGSOHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 378271

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/prodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/prodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	59.7	87	6	US-10-155-881-12211
2	34	54.8	93	5	US-09-547-599C-2428
3	34	54.8	96	6	US-10-143-775-583
4	33	53.2	76	5	US-09-540-209B-9470
5	33	53.2	97	6	US-10-217-648-645
6	32	51.6	15	5	US-09-350-401A-3025
7	32	51.6	15	5	US-09-350-401A-3183
8	32	51.6	61	6	US-10-106-698-6673
9	32	51.6	73	6	US-10-235-953-3603
10	32	51.6	89	5	US-09-513-986A-3966
11	31	50.0	37	6	US-10-105-299-5379
12	31	50.0	44	6	US-10-182-998-14849
13	31	50.0	44	6	US-10-203-135-30068
14	31	50.0	44	6	US-10-203-138-15288
15	31	50.0	44	6	US-10-203-139-30215
16	31	50.0	50	6	US-10-182-995-23617
17	31	50.0	50	6	US-10-203-134-29975
18	31	50.0	50	6	US-10-203-136-29938
19	31	50.0	50	6	US-10-182-993-29373
20	31	50.0	50	6	US-10-182-997-22308
21	31	50.0	50	6	US-10-182-998-13901
22	31	50.0	50	6	US-10-203-137-30270
23	31	50.0	50	6	US-10-203-138-14266
24	31	50.0	52	6	US-10-106-698-6662
25	31	50.0	57	6	US-10-106-698-6616
26	31	50.0	57	6	US-10-235-953-5210

27	31	50.0	64	6	US-10-212-759-1554	Sequence 1554, Ap
28	31	50.0	69	6	US-10-212-778-1146	Sequence 1146, Ap
29	31	50.0	71	6	US-10-182-995-24943	Sequence 24943, A
30	31	50.0	71	6	US-10-203-134-31357	Sequence 31357, A
31	31	50.0	71	6	US-10-203-136-31414	Sequence 31414, A
32	31	50.0	71	6	US-10-182-993-30655	Sequence 30655, A
33	31	50.0	71	6	US-10-182-997-23638	Sequence 23638, A
34	31	50.0	71	6	US-10-203-135-30509	Sequence 30509, A
35	31	50.0	71	6	US-10-203-137-31597	Sequence 31597, A
36	31	50.0	71	6	US-10-203-139-30553	Sequence 30553, A
37	31	50.0	79	5	US-09-791-537-111453	Sequence 111453, A
38	31	50.0	82	1	PCT-US02-09921-833	Sequence 833, App
39	31	50.0	84	5	US-09-684-197A-5	Sequence 5, App11
40	31	50.0	90	6	US-10-182-993-37688	Sequence 37688, A
41	31	50.0	90	6	US-10-203-139-37109	Sequence 37109, A
42	31	50.0	95	6	US-10-217-607-1183	Sequence 1183, Ap
43	30	48.4	57	4	US-08-905-054B-26	Sequence 26, Appl
44	30	48.4	69	6	US-10-235-954-752	Sequence 752, App
45	30	48.4	71	6	US-10-040-862-1215	Sequence 1215, Ap

#### ALIGNMENTS

RESULT 1  
US-10-155-881-12211  
Sequence 12211, Application US/10155881  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovacic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
FILE REFERENCE: 38-21(15300)J  
CURRENT APPLICATION NUMBER: US/10/155, 881  
CURRENT FILING DATE: 2002-05-22  
NUMBER OF SEQ. ID NOS: 37595  
SEQ ID NO 12211  
LENGTH: 87  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-155-881-12211

Query Match 59.7%, Score 37, DB 6, Length 87,  
Best Local Similarity 41.7%, Pred. No. 32,  
Matches 5, Conservative 5, Mismatches 2, Indels 0, Gaps 0;  
Db 42 LEVPGAHIDE 53

RESULT 2  
US-09-547-599C-2428  
Sequence 2428, Application US/09547599C  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Duclert, Aymeric  
APPLICANT: Lacroix, Bruno  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins  
FILE REFERENCE: GEN-711301  
CURRENT APPLICATION NUMBER: US/09/547, 599C  
CURRENT FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: US 08/905, 223  
PRIOR FILING DATE: 1997-08-01  
PRIOR APPLICATION NUMBER: US 08/905, 135  
PRIOR FILING DATE: 1997-08-01  
PRIOR APPLICATION NUMBER: US 08/905, 051  
PRIOR FILING DATE: 1997-08-01  
PRIOR APPLICATION NUMBER: US 08/905, 144  
PRIOR FILING DATE: 1997-08-01

```

; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: Patent.pm
; SEQ ID NO 2428
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Normal prostate
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -60..-1
; OTHER INFORMATION: Von Heijne matrix
; FEATURE:
; OTHER INFORMATION: score 7.00
; FEATURE:
; OTHER INFORMATION: seq VLVVLAIRSIGRS/CS
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 30
; OTHER INFORMATION: Xaa = *,Tyr
; US-09-547-599C-2428

Query Match
Best Local Similarity 54.8%; Score 34; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PGSQHI 9
Db 81 PGSQHI 86

RESULT 3
US-10-143-775-583
; Sequence 583, Application US/10143775
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC02ICIN
; CURRENT APPLICATION NUMBER: US/10/143,775
; CURRENT FILING DATE: 2002-05-14
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 1064
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 583
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-143-775-583

Query Match
Best Local Similarity 54.8%; Score 34; DB 6; Length 96;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 PGSQHI 10
Db 76 PGSQHI 82

RESULT 4
US-09-540-209B-9470
; Sequence 9470, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; TITLE OF INVENTION: PCR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
```

```

; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 9470
; LENGTH: 76
; TYPE: PRT
; ORGANISM: B.fragilis
; US-09-540-209B-9470

Query Match
Best Local Similarity 53.2%; Score 33; DB 5; Length 76;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYPSQHI 10
Db 34 EYPSQHI 42

RESULT 5
US-10-217-648-645
; Sequence 645, Application US/10217648
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM015CIN
; CURRENT APPLICATION NUMBER: US/10/217,648
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 09/758,440
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 832
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 645
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (64)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-217-648-645

Query Match
Best Local Similarity 53.2%; Score 33; DB 6; Length 97;
Best Local Similarity 45.5%; Pred. No. 2e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYPSQHI 12
Db 31 EYPSQHI 41

RESULT 6
US-09-350-401A-3025
; Sequence 3025, Application US/09350401A
; GENERAL INFORMATION:
; APPLICANT: Alessandro Sette
; APPLICANT: John Sidney
; APPLICANT: Scott Southwood
; APPLICANT: Maria A. Vitiello
; APPLICANT: Brian D. Livingston
; APPLICANT: Esteban Celis
; APPLICANT: Ralph T. Kubo
; APPLICANT: Howard M. Grey
; APPLICANT: Robert Chesnut
; TITLE OF INVENTION: INDUCING CELLULAR IMMUNE RESPONSES TO
; TITLE OF INVENTION: HEPATITIS B VIRUS USING PEPTIDE AND NUCLEIC ACID
```

```
; TITLE OF INVENTION: COMPOSITIONS
; FILE REFERENCE: 2060.0060008
; CURRENT APPLICATION NUMBER: US/09/350,401A
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: US 08/820,360
; PRIOR FILING DATE: 1997-03-12
; PRIOR APPLICATION NUMBER: US 60/013,363
; PRIOR FILING DATE: 1996-03-13
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-07-04
; NUMBER OF SEQ ID NOS: 3877
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3025
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Peptide
US-09-350-401A-3025

Query Match          51.6%; Score 32; DB 5; Length 15;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 VEPGSGHID 10
        || || || |
Db      6 VEPGSGHID 15

RESULT 7
US-09-350-401A-3183
; Sequence 3183, Application US/09350401A
; GENERAL INFORMATION:
; APPLICANT: Alessandro Sette
; APPLICANT: John Sidney
; APPLICANT: Scott Southwood
; APPLICANT: Maria A. Vitello
; APPLICANT: Brian D. Livingston
; APPLICANT: Esteban Cells
; APPLICANT: Ralph T. Kudo
; APPLICANT: Howard M. Grey
; APPLICANT: Robert Chesnut
; TITLE OF INVENTION: INDUCING CELLULAR IMMUNE RESPONSES TO
; TITLE OF INVENTION: HEPATITIS B VIRUS USING PEPTIDE AND NUCLEIC ACID
; TITLE OF INVENTION: COMPOSITIONS
; FILE REFERENCE: 2060.0060008
; CURRENT APPLICATION NUMBER: US/09/350,401A
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: US 08/820,360
; PRIOR FILING DATE: 1997-03-12
; PRIOR APPLICATION NUMBER: US 60/013,363
; PRIOR FILING DATE: 1996-03-13
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-07-04
; NUMBER OF SEQ ID NOS: 3877
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3183
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Artificially Synthesized Peptide
US-09-350-401A-3183

Query Match          51.6%; Score 32; DB 5; Length 15;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 VEPGSGHID 10
        || || || |
Db      6 VEPGSGHID 15

RESULT 8
US-10-106-698-6673
; Sequence 6673, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6673
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6673

Query Match          51.6%; Score 32; DB 6; Length 61;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy      3 VPGSGHDSQ 12
        :||| | :|
Db      23 IPGSTRHASAQ 32

RESULT 9
US-10-235-953-3603
; Sequence 3603, Application US/10235953
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PH002C1N
; CURRENT APPLICATION NUMBER: US/10/235,953
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/758,471
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 5386
```

```
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 3603
;; LENGTH: 73
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (16)
;; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: misc_feature
;; LOCATION: (71)
;; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-235-953-3603

Query Match
Best Local Similarity 51.6%; Score 32; DB 6; Length 73;
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EVPGSOHDSQ 12
| | | | |
Db 18 EPPGSTHAGSQ 28

RESULT 10
US-09-513-996A-3966
; Sequence 3966, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2/50-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 3966
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..89 /
; OTHER INFORMATION: any n or xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..89 / Ceres Seq. ID 1016527
US-09-513-996A-3966

Query Match
Best Local Similarity 51.6%; Score 32; DB 5; Length 89;
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 VEPPGSOHDSQ 12
| | | | |
Db 30 VETPGCVHYKTE 41

RESULT 11
US-10-105-299-5379
; Sequence 5379, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5379
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-299-5379
```

```
Query Match
Best Local Similarity 50.0%; Score 31; DB 6; Length 37;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 6 SQHDSQ 12
| | | | |
Db 16 SQHDSQ 22

RESULT 12
US-10-182-998-14849
; Sequence 14849, Application US/10182998
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: PB 0004 WO 9
; CURRENT APPLICATION NUMBER: US/10/182,998
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15009
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 14849
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL033532.12
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P34274, EVALU 1.60e-02
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: BF574321.1, EVALU 1.00e-09
US-10-182-998-14849

Query Match
Best Local Similarity 50.0%; Score 31; DB 6; Length 44;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 VPQSOHDSQ 12
| | | | |
Db 2 VPTTDHIDE 11

RESULT 13
US-10-203-135-30068
; Sequence 30068, Application US/10203135
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
```



```
;; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN LUNG
;; FILE REFERENCE: PB 0004 WO 5
;; CURRENT APPLICATION NUMBER: US/10/203.135
;; CURRENT FILING DATE: 2002-08-02
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 26 May 2000 (26.05.00)
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 03 August 2000 (03.08.00)
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 03 October 2000 (03.10.00)
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 27 September 2000 (27.09.00)
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; NUMBER OF SEQ ID NOS: 37012
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine
;; SEQ ID NO 30068
;; LENGTH: 44
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL033532.12
;; FEATURE:
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;; FEATURE:
;; OTHER INFORMATION: SWISSPROT HIT: P34274, EVALUE 1.60e-02
;; FEATURE:
;; OTHER INFORMATION: EST_HUMAN HIT: BF574321.1, EVALUE 1.00e-09
US-10-203-135-30068
```

```
Query Match          50.0%; Score 31; DB 6; Length 44;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 VPGSOHDSQ 12
   ||:||||:
Db 2 VPTDHDITE 11
```

```
RESULT 14
US-10-203-138-15288
;; Sequence 15288, Application US/10203138
;; GENERAL INFORMATION:
;; APPLICANT: Molecular Dynamics, Inc.
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: PB 0004 WO 8
;; CURRENT APPLICATION NUMBER: US/10/203.138
;; CURRENT FILING DATE: 2002-08-02
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 26 May 2000 (26.05.00)
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 03 August 2000 (03.08.00)
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 03 October 2000 (03.10.00)
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 27 September 2000 (27.09.00)
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; NUMBER OF SEQ ID NOS: 15438
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine
```

```
;; SEQ ID NO 15288
;; LENGTH: 44
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL033532.12
;; FEATURE:
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
;; FEATURE:
;; OTHER INFORMATION: SWISSPROT HIT: P34274, EVALUE 1.60e-02
;; FEATURE:
;; OTHER INFORMATION: EST_HUMAN HIT: BF574321.1, EVALUE 1.00e-09
US-10-203-138-15288
```

```
Query Match          50.0%; Score 31; DB 6; Length 44;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 VPGSOHDSQ 12
   ||:||||:
Db 2 VPTDHDITE 11
```

```
RESULT 15
US-10-203-139-30215
;; Sequence 30215, Application US/10203139
;; GENERAL INFORMATION:
;; APPLICANT: Molecular Dynamics, Inc.
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;; FILE REFERENCE: PB 0004 WO 4
;; CURRENT APPLICATION NUMBER: US/10/203.139
;; CURRENT FILING DATE: 2002-08-02
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 26 May 2000 (26.05.00)
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 03 August 2000 (03.08.00)
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 03 October 2000 (03.10.00)
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 27 September 2000 (27.09.00)
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; NUMBER OF SEQ ID NOS: 37156
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine
;; SEQ ID NO 30215
;; LENGTH: 44
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL033532.12
;; FEATURE:
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.88
;; FEATURE:
;; OTHER INFORMATION: SWISSPROT HIT: P34274, EVALUE 1.60e-02
;; FEATURE:
;; OTHER INFORMATION: EST_HUMAN HIT: BF574321.1, EVALUE 1.00e-09
US-10-203-139-30215
```

```
Query Match          50.0%; Score 31; DB 6; Length 44;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 VPGSOHDSQ 12
   ||:||||:
Db 2 VPTDHDITE 11
```

Fri Oct 25 17:50:59 2002

us-09-786-648-3.lenlim.rapn

Page 6

Db 2 VPTDHDTE 11

Search completed: October 24, 2002, 15:47:34  
Job time : 27.1639 secs

---

us-09-786-648-3.lenlim.rpr

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

in search, using sw model

October 24, 2002, 15:31:03 ; Search time 9.04918 Seconds  
(without alignments)  
127.423 Million cell updates/sec

VEVPGSQHDSQ 12

LOSUM62  
Gapop 10.0 , Gapext 0.5

```
883138 segs, 96089334 residues
fits satisfying chosen parameters: 37652
```

length: 0  
length: 100

Minimum	Match	0%
Maximum	Match	100%
Listing	first 45 summaries	

pir1:\*  
 pir1:\*  
 pir2:\*  
 pir3:\*  
 pir4:\*

nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
 A:Title: Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*.  
 A:Reference number: A95000; M0ID:21357209; PMID:11463916

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	33	53.2	79	2	C95152	IS66 family element
2	33	53.2	79	2	H95167	IS66 family element
3	31.5	50.8	74	2	G683593	hypothetical prote
4	31	50.0	35	2	T02997	sigma-E factor reg
5	31	50.0	65	2	T24159	hypothetical prote
6	31	50.0	79	2	B980334	degenerate transpo
7	30	48.4	31	2	A332384	G-alpha-1 protein
8	30	48.4	70	2	I390588	hypothetical prote
9	30	48.4	84	2	T17876	hypothetical prote
10	30	48.4	90	2	E71527	hypothetical prote
11	30	48.4	98	2	S06461	dystrophin, brain
12	29	46.8	33	2	A60222	myelin basic prote
13	29	46.8	52	4	S70252	hypothetical prote
14	29	46.8	59	2	S15939	acetylactate synth
15	29	46.8	72	2	D98143	hypothetical prote
16	29	46.8	74	2	S72758	BL496_C2.182 prote
17	29	46.8	79	2	S77831	hypothetical prote
18	29	46.8	93	2	T20227	hypothetical prote
19	29	46.8	98	1	FEMW	hypothetical prote
20	29	46.8	98	1	JAU098	ferredoxin [2Fe-2S
21	28.5	46.0	40	2	F45095	phosphosystem I ligh
22	28	45.2	44	2	S51064	ribosomal protein
23	28	45.2	44	2	S40297	phosphoenolpyruvat
24	28	45.2	55	2	B27397	Na+/K+-exchangin
25	28	45.2	73	2	F83824	hypothetical prote
26	28	45.2	79	2	H64643	hypothetical prote
27	28	45.2	87	2	T06281	reverse transcript
28	28	45.2	90	2	D81684	conserved hypotet
29	28	45.2	90	2	C82815	hypothetical prote

30	28	45.2	96	2	AF3175	attachment protein
31	28	45.2	97	2	AH2175	hypothetical protein
32	28	45.2	97	2	AH3516	hypothetical protein
33	27.5	44.4	60	2	E83906	hypothetical protein
34	27	43.5	13	2	A54326	granular kallikrein
35	27	43.5	42	2	E97915	choline binding protein
36	27	43.5	57	2	S62869	non-histone protein
37	27	43.5	75	2	S25719	protein ECL - rat
38	27	43.5	77	2	T01660	probable phosphoglycerate kinase
39	27	43.5	79	2	C84077	nitrogen fixation
40	27	43.5	83	2	AD0081	hypothetical protein
41	27	43.5	84	2	D83940	hypothetical protein
42	27	43.5	87	2	T01665	tat protein - human
43	27	43.5	93	2	S60801	M protein precursor
44	27	43.5	97	2	S31789	apolipoprotein C2
45	27	43.5	97	2	A91283	hypothetical protein

## ALIGNMENTS

```

RESULT 1
C95152
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: C95152
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapl
nson, T.; Hickey, E.R.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75412.1; PID:g14972794; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI314

Query Match      53.2%  Score 33;  DB 2;  Length 79;
Best Local Similarity 50.0%  Pred. No. 39;
Matches 6;  Conservative 3;  Mismatches 3;  Indels 0;  Gaps 0;

OY      1  VEPGSOHDSO 12
      ||| | | | |
Db      26  VEIPOSRRFSK 37

RESULT 2
H95167
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95167
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapl
nson, T.; Hickey, E.R.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95167
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: PIDN:AAK75537.1; PID:g14972931; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI443

```

Query Match 53.2%; Score 33; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 39;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEYPSQHDSQ 12  
||:|:|:|:  
DB 26 VETPSRPFDSK 37

## RESULT 3

hypothetical protein PA2021 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: G83393

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bz  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lam,  
J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337

A:Accession: G83393

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-74 <STO>  
A:Cross-references: GB:AE004628; GB:AE004091; NID:g9948018; PIDN:AAG05409.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

Query Match 50.8%; Score 31.5; DB 2; Length 74;  
Best Local Similarity 50.0%; Pred. No. 70;  
Matches 7; Conservative 3; Mismatches 1; Indels 3; Gaps 1;  
QY 1 VEYPSQHDSQ 11  
||:|:|:|:  
DB 14 VETGSRHAPYDS 27

## RESULT 4

sigma-E factor regulatory protein rseA - Salmonella typhimurium (fragment)  
C:Species: Salmonella typhimurium

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999

C:Accession: T02997

R:Figuera-Bossi, N.; Bossi, L.  
submitted to the EMBL Data Library, June 1998

A:Reference number: Z14818

A:Accession: T02997

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-35 <FIG>

A:Cross-references: EMBL:AF001386; NID:g3294471; PIDN:AAC26063.1; PID:g3294474

A:Experimental source: strain LT2

C:Genetics:

Query Match 50.0%; Score 31; DB 2; Length 35;  
Best Local Similarity 50.0%; Pred. No. 38;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 VEYPSQHDSQ 12  
||:|:|:|:  
DB 22 VQVPGIGTGTQ 33

## RESULT 5

hypothetical protein R11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

R:McMurray, A.  
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19847

A:Accession: T24159

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-65 <WIL>

A:Cross-references: EMBL:Z81577; PIDN:CAB04648.1; GSPDB:GN00028; CESP:R11.4

A:Experimental source: clone R11

C:Genetics:

A:Gene: CESP:R11.4

A:Map position: X

Query Match 50.0%; Score 31; DB 2; Length 65;  
Best Local Similarity 50.0%; Pred. No. 76;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEYPSQHDSQ 12  
||:|:|:|:  
DB 21 VEMPAQSTIQEQ 32

RESULT 6  
B98034  
degenerate transposase (orf1) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: B98034

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
Y., P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98034

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <RUR>

A:Cross-references: GB:AE007317; PIDN:AAL00103.1; PID:g15458942; GSPDB:GN00174

C:Genetics:

A:Gene: IS1381-truncation

Query Match 50.0%; Score 31; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 94;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEYPSQHDSQ 12  
||:|:|:|:  
DB 26 VETPSRPFDSK 37

RESULT 7  
A32384  
G-alpha-1 protein - slime mold (Dictyostelium sp.) (fragment)  
C:Species: Dictyostelium sp.

C:Date: 08-Sep-1989 #sequence\_revision 08-Sep-1989 #text\_change 19-Jul-1996

C:Accession: A32384

R:Kumagai, A.; Pupillo, M.; Gundersen, R.; Maize-Lye, R.; Devreotes, P.N.; Firtel, R.  
Cell 57, 265-275, 1989

A:Title: Regulation and function of G-alpha protein subunits in Dictyostelium.

A:Reference number: A32384; MUID:89195237

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-31 <KUM>

C:Superfamily: GTP-binding regulatory protein Gs alpha chain

Query Match 48.4%; Score 30; DB 2; Length 31;  
Best Local Similarity 50.0%; Pred. No. 52;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVYPSQHIDS 11



C:Accession: S70252  
 R:Schnaubeier, B.; Neveling, U.; Schmitz, S.  
 M:1. Microbiol. 18, 357-369, 1995  
 A:Title: Distinct and differentially regulated Mo-dependent nitrogen-fixing systems evolved  
 regions as part of the nifH/2 gene clusters.  
 A:Reference number: S70242; MID:96296457  
 A:Accession: S70252  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-52 <SCH>  
 A:Cross-references: EMBL:U25172; NID:1160349  
 A:Experimental source: ATCC 29413  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
 C:Comment: This is the hypothetical translation of a sequence translated in an incorrect

Query Match 46.8%; Score 29; DB 4; Length 52;  
 Best Local Similarity 62.5%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEVPGSQH 8  
 | | | | |  
 Db 44 VSVPGTTPH 51

## RESULT 14

S15939  
 acetolactate synthase (EC 4.1.3.18) III large chain - Salmonella typhimurium (fragment)  
 C:Species: Salmonella typhimurium  
 A:Variety: strain LT2  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 15-Oct-1999  
 C:Accession: S15939; S14545  
 R:Jahreis, K.; Postma, P.W.; Lengeler, J.W.  
 M:1. Gen. Genet. 226, 332-336, 1991  
 A:Title: Nucleotide sequence of the ilvH-fruR gene region of Escherichia coli K12 and SA  
 A:Reference number: S15939; MID:91238714  
 A:Accession: S15939  
 A:Molecule type: DNA  
 A:Residues: 1-59 <JAH>  
 A:Cross-references: EMBL:X55456; NID:947690; PID:CAA39101.1; PID:947691  
 C:Genetics:  
 A:Gene: ilvH  
 C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
 C:Keywords: branched-chain amino acid biosynthesis; carbon-carbon lyase; flavoprotein; m

Query Match 46.8%; Score 29; DB 2; Length 59;  
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEVPGSQH 9  
 | | | | |  
 Db 30 VTVDGSEHV 38

## RESULT 15

D98143  
 hypothetical protein AGR\_L\_207 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
 C:Accession: D98143  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirrollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A:Reference number: A97359; PMID:11743194  
 A:Accession: D98143  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-72 <KUR>  
 A:Cross-references: GB:AE007870; PIDN:AAK88670.1; PID:q15158399; GSPDB:GN00170  
 C:Genetics:  
 A:Gene: AGR\_L\_207  
 A:Map position: linear chromosome

Query Match 46.8%; Score 29; DB 2; Length 72;  
 Best Local Similarity 57.1%; Pred. No. 2e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 PGSQHID 10  
 | | | | |  
 Db 64 PGSSHAVN 70

Search completed: October 24, 2002, 15:37:43  
 Job time : 11.0492 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:23:28 ; Search time 4.32787 Seconds  
(without alignments)  
107.359 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62  
Sequence: 1 VEVPGSQHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 12261

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	46.8	98	1	FER_MASIA
2	28	45.2	24	1	RS13_THETH
3	28	45.2	91	1	Y13F_BPT4
4	27.5	44.4	15	1	PH3_PPARAM
5	27.5	44.4	17	1	PH3_PPARAM
6	27.5	44.4	83	1	IMB8_MOUSE
7	27	43.5	67	1	CSPA_BACCE
8	27	43.5	76	1	UBIQ_DICDI
9	27	43.5	87	1	TAT_HVIMA
10	27	43.5	90	1	2600_DROME
11	27	43.5	97	1	APC2_MOUSE
12	27	43.5	100	1	RL23_MYCLE
13	27	43.5	100	1	YABP_BACSU
14	26	41.9	29	1	GLUC_CHIBR
15	26	41.9	74	1	PURF_RHOCA
16	26	41.9	82	1	DNM1_STRAL
17	26	41.9	88	1	LCR8_YERPE
18	26	41.9	92	1	BM6_ARATH
19	26	41.9	94	1	RL25_ECOLI
20	26	41.9	97	1	FER2_CYACA
21	26	41.9	98	1	FER1_ANASP
22	26	41.9	98	1	FER2_NOSMU
23	26	41.9	98	1	FER_NOSMU
24	26	41.9	98	1	FER_STNPA
25	26	41.9	99	1	FER1_PLEBO
26	26	41.9	99	1	FLGM_YEREN
27	26	41.9	100	1	RL23_MYCBO
28	26	41.9	100	1	RL23_MYCTU
29	25	40.3	43	1	Y123_HELIP
30	25	40.3	64	1	SCX3_MESMA
31	25	40.3	64	1	SCX3_LEIOU
32	25	40.3	64	1	SCX4_MESMA
33	25	40.3	76	1	CD24_MOUSE

34	25	40.3	79	1	SCX9_MESMA	P45698 mesobuthus
35	25	40.3	80	1	ELB8_KLEPN	Q48464 Klebsiella
36	25	40.3	81	1	HPIS_THRPF	P00263 thiocapsa p
37	25	40.3	84	1	SC13_MESMA	Q9njc8 mesobuthus
38	25	40.3	84	1	SC16_MESMA	Q9qy66 mesobuthus
39	25	40.3	84	1	SC16_MESMA	P45697 mesobuthus
40	25	40.3	84	1	SC16_MESMA	Q9n682 mesobuthus
41	25	40.3	84	1	YCAF_ECOLI	P76244 escherichia
42	25	40.3	85	1	SC12_MESMA	Q9qy43 mesobuthus
43	25	40.3	85	1	SCX4_LEIOU	P17728 leiurus qui
44	25	40.3	85	1	Y9KD_STRLI	P22399 streptomyce
45	25	40.3	88	1	CH10_TREPA	O83977 treponema p

## ALIGNMENTS

```

RESULT 1
FER_MASIA STANDARD; PRT; 98 AA.
AC P00248; O31125;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ferredoxin.
GN PETF.
OS Mastigocladus laminosus (Fischerella sp.).
OC Bacteria; Cyanobacteria; Stigonematales; Fischerella.
OX NCBI_TaxID=1191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC 7605;
RA He Z.-Y., Chittus P.R., Nechushtai R.;
RT "Molecular cloning of the petf gene encoding ferredoxin I of the
RL thermophilic cyanobacterium Mastigocladus laminosus.";
RN [2]
RP SEQUENCE.
RC STRAIN=COHN.
RA Hase T., Wakabayashi S., Matsubara H., Rao K.K., Hall D.O.;
RL Submitted (SEP-1978) to the PIR data bank.
CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -1- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF030002; AAC04840.1; -.
DR PIR; A00252; FEWM.
DR HSSP; P06543; ICZP.
DR InterPro; IPR000564; 2Fe2S-ferredoxin.
DR InterPro; IPR001041; Ferredoxin.
DR Pfam; PF00111; fer2; 1.
DR PRINTS; PR00159; 2FE2SFDOXIN.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur.
FT INITI-MET 0
FT METAL 41
FT METAL 41
FT METAL 46
FT METAL 49
FT METAL 79
SQ SEQUENCE 98 AA; 10617 MW; 1D6AF4702ACDE707 CAC64;
Query Match 46.8%; Score 29; DB 1; Length 98;
Best Local Similarity 55.6%; Prod. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 VEPGSOHI 9  
 :||| :|||  
 Db 18 IEVDDOXY 26

## RESULT 2

RS13\_THETH STANDARD: PRT: 24 AA.  
 AC P80377;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE 30S ribosomal protein S13 (Fragment).  
 GN RPSM OR RPS13.  
 OS Thermus aquaticus (subsp. thermophilus).  
 OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.  
 OX NCBI\_TaxID=274;

RP SEQUENCE.  
 RX MEDLINE=95045586; PubMed=7957245;  
 RA Tshobli P., Herfurth E., Choll T.;  
 RT "Purification and characterization of the 30S ribosomal proteins from  
 the bacterium Thermus thermophilus."  
 RL Eur. J. Biochem. 226:169-177(1994).  
 CC -1- FUNCTION: INVOLVED IN THE BINDING OF EMET-TRNA AND, HENCE, IN THE  
 INITIATION OF TRANSLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.  
 DR InterPro: IPR001892; RIBOSOMAL\_S13.  
 DR Pfam: PF00416; RIBOSOMAL\_S13; PARTIAL.  
 KW PROSITE, PS00646; RIBOSOMAL\_S13; PARTIAL.  
 FT RIBOSOMAL protein.  
 SQ SEQUENCE 24 AA; 2688 MW; 337F88EA0294D2A8 CRC64;

Query Match 45.2%; Score 28; DB 1; Length 24;  
 Best Local Similarity 40.0%; Pred. No. 33;  
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHI 10  
 :||| :|||  
 Db 6 VEIPKRRVD 15

## RESULT 3

Y13F\_BPPT4 STANDARD: PRT: 91 AA.  
 ID Y13F\_BPPT4  
 AC P39499;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 10.1 kDa protein in cd-psrf intergenic region.  
 GN Y13F OR CD.3.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like phages.  
 OX NCBI\_TaxID=10665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuter E., Arisaka F., Kunitawa T., Tsugita A., Mosig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis";  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF158101; AAD42549.1; -

KW Hypothetical protein.  
 SQ SEQUENCE 91 AA; 10131 MW; B0010922E122A21F CRC64;

Query Match 45.2%; Score 28; DB 1; Length 91;  
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 PGSOHI 9  
 :||| :|||  
 Db 58 PGSAMI 63

## RESULT 4

PH2\_PERAM STANDARD: PRT: 15 AA.  
 ID PH2\_PERAM  
 AC P82695;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Peptide hormone 2 (Pea-VEAcid 2).  
 OS Periplaneta americana (American cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattodea; Blattidae; Periplaneta.  
 OX NCBI\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Abdominal perisymphathetic organs;  
 RA Predel R.;  
 RL Submitted (JUL-2000) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: UNKNOWN.  
 KW Neuropeptide.  
 SQ SEQUENCE 15 AA; 1603 MW; F353DC8B1F92B8BD CRC64;

Query Match 44.4%; Score 27.5; DB 1; Length 15;  
 Best Local Similarity 75.0%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 4 PGSOHI 11  
 :||| :|||  
 Db 5 PGS-HVDS 11

## RESULT 5

PH3\_PERAM STANDARD: PRT: 17 AA.  
 ID PH3\_PERAM  
 AC P82696;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Peptide hormone 3 (Pea-VEAcid 1).  
 OS Periplaneta americana (American cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattodea; Blattidae; Periplaneta.  
 OX NCBI\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Abdominal perisymphathetic organs;  
 RX MEDLINE=20140865; PubMed=10676456;  
 RA Predel R., Eckert M., Holman G.M.;  
 RT "The unique neuropeptide pattern in abdominal perisymphathetic organs  
 of insects";  
 RL Ann. N.Y. Acad. Sci. 897:282-290(1999).  
 CC -1- FUNCTION: UNKNOWN.  
 CC -1- MASS SPECTROMETRY: MW=1651.35; METHOD=MALDI.  
 KW Neuropeptide.  
 SQ SEQUENCE 17 AA; 1807 MW; 237AC8B1F86EBB CRC64;

Query Match 44.4%; Score 27.5; DB 1; Length 17;  
 Best Local Similarity 75.0%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 1;



OY 4 PGSQHDS 11  
 DB 7 PGS-HVDS 13

# RESULT 6

ID IM8\_MOUSE STANDARD; PRT; 83 AA.

AC 09OCT4; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mitochondrial Import inner membrane translocase subunit TIM8 B

GN (Deafness dystonia protein 2 homolog).

OS TIM8B OR TIM8B OR DDP2.

OS Mus musculus (Mouse), and

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090, 10116;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Mouse, and Rat;

RA Bauer M.F., Brunner M., Hofmann S.;

RT Cloning and mapping of Tim10/DDP gene family encoding small zinc

RT finger proteins involved in mitochondrial carrier import.

RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By

similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF196315; AAF13329.1; -

DR EMBL: AF196314; AAF13328.1; -

DR MGD; MGI:1353424; TImmbd.

DR InterPro: IPR004217; zif-Tim10-DDP.

DR Pfam: PF02953; zif-Tim10-DDP; 1.

KW Transport; Protein transport; Translocation; Mitochondrion;

KW Inner membrane.

SO SEQUENCE 83 AA; 9286 MW; 04C47BA531CB689C CRC64;

OY 1 VEVPGSQHDSQ 12

DB 41 VEPGSR-LDSR 51

Query Match 44.4%; Score 27.5; DB 1; Length 83;

Best Local Similarity 58.3%; Pred. No. 1.6e+02;

Matches 7; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

RESULT 7

ID CSPA\_BACE STANDARD; PRT; 67 AA.

AC 04OCT4; 16-OCT-2001 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE 15-JUL-1998 (Rel. 36, Last annotation update)

GN Major cold shock protein cspa.

OS CSPA.

OC Bacillus cereus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1396;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-28.

RC STRAIN=MSBC 10201;

RX MEDLINE=96213034; PubMed=8631682;

RA Mayr B., Kaplan T., Lechner S., Scherer S.;

RT Identification and purification of a family of dimeric major cold

RT shock protein homologs from the psychrotrophic *Bacillus cereus* MSBC

RT 10201.

RL J. Bacteriol. 178:2916-2925(1996).

CC -1- FUNCTION: CAN BIND TO ATGGS AND CCAAT MOTIFS (Y-BOX MOTIFS) OF

CC SINGLE STRANDED OLIGONUCLEOTIDES.

CC -1- SUBUNIT: HOMODIMER.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- INDUCTION: IN RESPONSE TO LOW TEMPERATURE.

CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X93039; CAA63607.1; -

DR HSSP; P41016; IC90.

DR InterPro: IPR02059; Cold\_shock.

DR Pfam; PF00313; CSD; 1.

DR PRINTS; PR00050; COLDSHOCK.

DR Prodom; PD000621; Cold\_shock; 1.

DR SMART; SM00357; CSP; 1.

DR PROSITE; PS00352; COLD\_SHOCK; 1.

KW Transcription regulation; DNA-binding; Activator; Multigene family.

FT DOMAIN 5 64

FT CONFLICT 23 23 G->GC (IN REF. 1; AA SEQUENCE).

SO SEQUENCE 67 AA; 7491 MW; 284EB4CDB554B73F CRC64;

OY 1 VEVPGSQH 9

DB 19 IVEPGENDV 27

Query Match 43.5%; Score 27; DB 1; Length 67;

Best Local Similarity 44.4%; Pred. No. 1.6e+02;

Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

RESULT 8

ID UBIO\_DICDI STANDARD; PRT; 76 AA.

AC P08618; 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DE 01-NOV-1991 (Rel. 20, Last annotation update)

DE Ubiquitin.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

OX NCBI\_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88152253; PubMed=2831095;

RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,

RA Gerisch G.;

RT "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-

RT terminal tail and identification of the protein using an anti-peptide

RT antibody";

RL FEBS Lett. 229:273-278(1988).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=89352609; PubMed=2548604;

RA Ohmachi T., Giorda R., Shaw D.R., Ennis H.L.;

RT "Molecular organization of developmentally regulated Dictyostelium

RT discoideum ubiquitin cDNAs";

RL Biochemistry 28:5226-5231(1989).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=87257921; PubMed=3037345;

RA Giorda R., Ennis H.L.;

```

RT "Structure of two developmentally regulated Dictyostellium discoideum
RT ubiquitin genes."
RT Mol. Cell. Biol. 7:2097-2103(1987).
RN [4]
RN SEQUENCE OF 13-76 FROM N.A.
RC STRAIN-AX2:
RA Westphal M., Mueller-Taubenberg A., Noegel A., Gerisch G.:
RT "Transcript regulation and carboxyterminal extension of ubiquitin in
RT Dictyostellium discoideum."
RT FEBS Lett. 209:92-96(1986).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE. THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES
CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07210; CAA30183.1; ALT_TERM.
DR EMBL: M19666; AAA33261.1; ALT_TERM.
DR EMBL: M19491; AAA33269.1; ALT_TERM.
DR EMBL: M19492; AAA33270.1; ALT_TERM.
DR EMBL: M23748; AAA33262.1; ALT_TERM.
DR EMBL: M23749; AAA33263.1; ALT_TERM.
DR EMBL: M23750; AAA33264.1; ALT_TERM.
DR EMBL: M23751; AAA33265.1; ALT_TERM.
DR EMBL: M23752; AAA33266.1; ALT_TERM.
DR EMBL: M23753; AAA33267.1; ALT_TERM.
DR EMBL: M23754; AAA33268.1; ALT_TERM.
DR EMBL: X04702; CAA28408.1; ALT_TERM.
DR PIR: S00357; UODOR.
DR PIR: E34080; UODOR7.
DR PIR: A27806; A27806.
DR PIR: B27806; B27806.
DR PIR: A34080; A34080.
DR PIR: B34080; B34080.
DR PIR: C34080; C34080.
DR HSSP: P02248; D34080.
DR DictyDB: DD05001; ubpA.
DR InterPro: IPR000626; ubiquitin.
DR Pfam: PF00240; ubiquitin.1.
DR PRINTS: PR00348; UBIQUITIN.
DR SMART: SM00213; UBO.1.
DR PROSITE: PS00299; UBIQUITIN.1; 1.
DR PROSITE: PS50053; UBIQUITIN.2; 1.
KW Nuclear protein; Polyprotein.
FT SITE 48 48
FT BINDING 76 76
FT VARIANT 11 11
FT SEQUENCE 76 AA; 8538 MW; 6427383968E8A84 CRC64;
SQ
Query Match 43.5%; Score 27; DB 1; Length 76;
Best Local Similarity 45.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 VEPGSOHDS 11
DB 15 LEVSGSDNIEN 25

```

```

ID TAT_HV1MA STANDARD: PRT; 87 AA.
AC P04613;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TAT protein (Transactivating regulatory protein).
GN TAT.
OS Human immunodeficiency virus type 1 (MAL isolate) (HIV-1).
OC Nucleus; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11697;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86245056; PubMed=2424612;
RA Allison M., Main-Hobson S., Montagnier L., Sonigo P.:
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis
RT of two isolates from African patients."
RL Cell 46:63-74(1986).
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE
CC TRANS-ACTIVATING RESPONSE SEQUENCE (TAR) RNA ELEMENT AND
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR
CC PROMOTER.
CC -1- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04415; CAA28015.1; -.
DR HIV: K03456; TATSMAL.
DR InterPro: IPR001831; HIV_Tat.
DR Pfam: PF00539; Tat.1.
DR PRINTS: PR00055; HIVTATDOMAIN.
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;
KW AIDS.
SQ SEQUENCE 87 AA; 10036 MW; 3832412849D5B1CE CRC64;
OY 2 EWPSSGH 8
DB 71 EWPSSGH 77

```

```

RESULT 10
ID 2600_DROME STANDARD: PRT; 90 AA.
AC P22469; O9VUP1;
DT 13-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein 2600.
GN 2600 OR CG17962.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232719; PubMed=2497054;
RA Schulz R.A., Butler B.A.:
RT "Overlapping genes of Drosophila melanogaster: organization of the
RT 2600-gonadal-Bip28/29 gene cluster."
RL Genes Dev. 3:232-242(1989).
RN [2]
SQ SEQUENCE FROM N.A.

```

RESULT 9  
TAT\_HV1MA

RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; Pubmed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burdits R.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwam C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: NOT KNOWN.  
 CC -1- DEVELOPMENTAL STAGE: TEMPORALLY RESTRICTED TO EARLY  
 CC EMBRYOGENESIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X58286; CAA41222.1; -;  
 DR EMBL; AE003530; AAF49634.1; -;  
 DR PIR; A30172; A30172.  
 DR FlyBase; FBgn0004052; Z600.  
 FT CONFLICT 27 27 H -> N (IN REF. 1).  
 FT SEQUENCE 90 AA; 10530 MW; 19054C5965EE9109 CRC64;  
 SQ  
 Query Match 43.5%; Score 27; DB 1; Length 90;  
 Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Apolipoprotein C-II precursor (Apo-CII).  
 GN APOC2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/RIJ / 129/J; TISSUE=Liver;  
 RX MEDLINE-93162679; Pubmed=7916738;  
 RA Hoffer M.J.V., Hoffer M.H., van Eck M.M., Havekes L.M.,  
 RA Frants R.R.;  
 RT "Evolutionary conservation of the mouse apolipoprotein E-C1-C2 gene  
 RT cluster: structure and genetic variability in inbred mice.";  
 RL Genomics 15:62-67(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/RIJ;  
 RX MEDLINE-94010908; Pubmed=7691714;  
 RA Hoffer M.J.V., van Eck M.M., Havekes L.M., Hoffer M.H.,  
 RA Frants R.R.;  
 RT "Structure and expression of the mouse apolipoprotein C2 gene.";  
 RL Genomics 17:45-51(1993).  
 CC -1- FUNCTION: APOC-II IS A COMPONENT OF THE VERY LOW DENSITY  
 CC LIPOPROTEIN (VLDL) FRACTION IN PLASMA, AND IS AN ACTIVATOR OF  
 CC SEVERAL TRIGLYCEROL LIPASES. THE ASSOCIATION OF APOC-II WITH  
 CC PLASMA CHYLOMICRONS, VLDL, AND HDL IS REVERSIBLE. A FUNCTION OF  
 CC THE SECRETION AND CATABOLISM OF TRIGLYCERIDE-RICH LIPOPROTEINS,  
 CC AND CHANGES RAPIDLY.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Secreted in plasma.  
 CC -1- SIMILARITY: BELONGS TO THE APOC2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z15090; CAA78804.1; -;  
 DR EMBL; Z22216; CAA80220.1; -;  
 DR EMBL; Z22217; CAA80220.1; JOINED.  
 DR PIR; S31799; S31799.  
 DR MGD; MGI:88054; APOC2.  
 KW Chylomicron; VLDL; Plasma; Lipid transport; Lipid degradation; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 97  
 FT BINDING 40 48  
 FT DOMAIN 52 75  
 FT FT  
 FT CONFLICT 89 89  
 FT SEQUENCE 97 AA; 10741 MW; 18223E9034A1983E CRC64;  
 SQ  
 Query Match 43.5%; Score 27; DB 1; Length 97;  
 Best Local Similarity 60.0%; Pred. No. 2.4e+02;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 11  
 APC2\_MOUSE STANDARD; PRT; 97 AA.  
 AC 005020;  
 DT 01-FEB-1994 (Rel. 28, Created)

RESULT 12  
 RL23\_MYCLE STANDARD; PRT; 100 AA.  
 AC 032983;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 50S ribosomal protein L23.  
 CC RPLM OR ML1861 OR MLCB2492.04.  
 OS Mycobacterium leprae.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 CC NCBI\_TaxID=1769;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=TN;  
 RC MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 408:1007-1011(2001).  
 CC -1- FUNCTION: BINDS TO A SPECIFIC REGION ON THE 23S RNA  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; 298756; CAB1464.1; -;  
 CC DR EMBL; AL583923; CAC30815.1; -;  
 CC DR Leproma; ML1861; -;  
 CC DR InterPro: IPR001014; Ribosomal\_L23.  
 CC DR Pfam: PF00276; Ribosomal\_L23; 1.  
 CC DR ProDom: PD001141; Ribosomal\_L23; 1.  
 CC DR PROSITE; PS00050; RIBOSOMAL\_L23; 1.  
 CC KW Ribosomal protein; rRNA-binding; Complete proteome.  
 CC SQ SEQUENCE 100 AA; 10939 MW; 1A9254F7D639D1AF CRC64;  
 CC  
 CC Query Match 43.5%; Score 27; DB 1; Length 100;  
 CC Best Local Similarity 71.4%; Pred. No. 2.5e+02;  
 CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC QY 4 PGSQHD 10  
 CC DB 88 PGSKSD 94  
 CC  
 CC RESULT 13  
 CC YABP\_BACSU STANDARD; PRT; 100 AA.  
 CC AC P37558;  
 CC DT 01-OCT-1994 (Rel. 30, Created)  
 CC DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Hypothetical 11.4 kDa protein in MFD-DIVIC intergenic region.  
 CC GN YABP.  
 CC OS Bacillus subtilis.  
 CC CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC CC Bacillus/Staphylococcus group; Bacillus.  
 CC OX NCBI\_TaxID=1423;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=168;  
 CC RC MEDLINE=96051385; PubMed=7584024;  
 CC RX Ogatawara N., Nakai S., Yoshikawa H.;  
 CC RA "Systematic sequencing of the 180 kilobase region of the Bacillus  
 CC RT subtilis chromosome containing the replication origin."  
 CC RL DNA Res. 1:1-14(1994).

CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; D26185; BAA05295.1; -;  
 CC DR EMBL; 299104; CAB1836.1; -;  
 CC DR Sublist: BG10123; yabp.  
 CC KW Hypothetical protein; Complete proteome.  
 CC SQ SEQUENCE 100 AA; 11403 MW; 0A3DFD6A75E8FF79 CRC64;  
 CC  
 CC Query Match 43.5%; Score 27; DB 1; Length 100;  
 CC Best Local Similarity 27.3%; Pred. No. 2.5e+02;  
 CC Matches 3; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
 CC  
 CC QY 1 VEPYSGHIDS 11  
 CC DB 27 LDISGVKVES 37  
 CC  
 CC RESULT 14  
 CC GLUC\_CHIBR STANDARD; PRT; 29 AA.  
 CC AC P31297;  
 CC DT 01-JUL-1993 (Rel. 26, Created)  
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 CC DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 CC DE Glucagon.  
 CC OS Chinchilla brevicauda (Chinchilla).  
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC CC Mammalia; Eutheria; Rodentia; Hystriocomnathi; Chinchillidae;  
 CC CC Chinchilla.  
 CC OX NCBI\_TaxID=10152;  
 CC RN [1]  
 CC RP SEQUENCE.  
 CC RX MEDLINE=91045327; PubMed=2235678;  
 CC RA Eng J., Kleinman W.A., Chu L.S.;  
 CC RT "Purification of peptide hormones from chinchilla pancreas by  
 CC RT chemical assay."  
 CC RL Peptides 11:683-685(1990).  
 CC CC -1- FUNCTION: PROMOTES HYDROLYSIS OF GLYCOGEN AND LIPIDS, AND RAISES  
 CC CC THE BLOOD SUGAR LEVEL.  
 CC CC -1- INDUCTION: PRODUCED IN THE A CELLS OF THE ISLETS OF LANGERHANS  
 CC CC IN RESPONSE TO A DROP IN BLOOD SUGAR CONCENTRATION.  
 CC CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.  
 CC CC PIR: A60413; GCCB.  
 CC DR HSSP; P01274; IGCN.  
 CC DR InterPro: IPR000532; Glucagon.  
 CC DR Pfam; PF00123; hormone2; 1.  
 CC DR PRINTS; PR00275; GLUCAGON.  
 CC DR SMART; SM00070; GLUCA; 1.  
 CC DR PROSITE; PS00260; GLUCAGON; 1.  
 CC KW Glucagon family; Hormone.  
 CC SQ SEQUENCE 29 AA; 3478 MW; 19ECF4DAB752B27 CRC64;  
 CC  
 CC Query Match 41.9%; Score 26; DB 1; Length 29;  
 CC Best Local Similarity 57.1%; Pred. No. 98;  
 CC Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 6 SQHSDS 12  
 CC DB 11 SKHDSR 17  
 CC  
 CC RESULT 15  
 CC PUFO\_RHOCA STANDARD; PRT; 74 AA.  
 CC ID PUFO\_RHOCA  
 CC AC P14601;  
 CC DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, last sequence update)  
 DE 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Protein pufo.  
 GN pufo.  
 OS Rhododactyl capsulatus (Rhodospirillum rubrum).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhododactyl group;  
 OC Rhododactyl.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8169603; PubMed=3127391;  
 RA Bauer C.E., Young D.A., Mairs B.L.;  
 RT "Analysis of the Rhododactyl capsulatus pufo operon. Location of the  
 RT oxygen-regulated promoter region and the identification of an  
 RT additional pufo-encoded gene.";  
 RL J. Biol. Chem. 263:4820-4827(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89123059; PubMed=2492501;  
 RA Adams C.W., Forrest M.E., Cohen S.N., Beatty J.T.;  
 RT "Structural and functional analysis of transcriptional control of the  
 RT Rhododactyl capsulatus pufo operon.";  
 RL J. Bacteriol. 171:473-482(1989).  
 CC -1- FUNCTION: REQUIRED FOR BACTERIOCHLOROPHYLL BIOSYNTHESIS. DIRECTLY  
 CC INVOLVED IN THE ASSEMBLY OF BOTH THE B875 AND B800-850 PIGMENT-  
 CC PROTEIN COMPLEXES.  
 CC -1- SIMILARITY: BELONGS TO THE PUFO FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 DR EMBL: M20141; AAA26169.1; -;  
 DR EMBL: M23752; AAA26168.1; -;  
 DR EMBL: Z11165; CAA77551.1; -;  
 DR PIR: A28143; A28143;  
 DR PIR: A32253; A32253;  
 KW Photosynthesis; Chlorophyll biosynthesis.  
 SQ SEQUENCE 74 AA; 8555 MW; 5504D8AFBE6923BB CRC64;

Query Match 41.9%; Score 26; DB 1; Length 74;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 5 GSOHID 10  
 | | | | |  
 Db 10 GVQHYD 15

Search completed: October 24, 2002, 15:35:31  
 Job time : 6.32787 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:24:04 : Search time 13.9672 Seconds  
(without alignments)  
148.629 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEPGSHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 107123

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Maximum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPTRMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.rodent:\*  
13: sp.virus:\*  
14: sp.vertibrate:\*  
15: sp.unclassified:\*  
16: sp.virus:\*  
17: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	59.7	79	4	096HT2
2	35	56.5	99	8	037764
3	33	53.2	77	11	090YJ8
4	33	53.2	79	16	097OB1
5	33	53.2	79	16	097PZ4
6	32	51.6	73	6	09GMA0
7	32	51.6	93	3	09U194
8	32	51.6	95	10	09FOI8
9	31.5	50.8	74	16	091293
10	31	50.0	35	2	084884
11	31	50.0	58	4	096HY2
12	31	50.0	65	5	062344
13	31	50.0	65	10	09M4H3
14	31	50.0	83	7	019274
15	31	50.0	87	4	015962
16	31	50.0	91	5	09BNH0

17	31	50.0	91	5	09BN97	09bn97 tachyurod
18	31	50.0	95	10	09FOJ8	09fj8 amarantus
19	30	48.4	27	13	090WP2	090wp2 fugu rubrip
20	30	48.4	47	12	091IL6	091il6 human metap
21	30	48.4	61	5	077288	077288 drosophila
22	30	48.4	64	11	092022	092022 rattus norv
23	30	48.4	80	10	09ZTSS	09zts5 zea mays (m
24	30	48.4	84	12	098428	098428 paramedum
25	30	48.4	90	11	P70443	P70443 mus musculu
26	30	48.4	90	16	084334	084334 chlamydia t
27	30	48.4	91	5	09BNC8	09bnc8 neacarus te
28	30	48.4	91	5	09BNC2	09bnc2 parasitid '
29	30	48.4	91	5	09BN47	09bn47 sejus sp. '
30	30	48.4	91	5	09BN94	09bn94 zircon sp.
31	30	48.4	93	2	09R349	09r349 streptomyce
32	30	48.4	99	11	063770	063770 rattus norv
33	29	46.8	67	11	025833	025833 rattus norv
34	29	46.8	73	5	025833	025833 mycobacteri
35	29	46.8	74	2	049687	049687 mycoplasma
36	29	46.8	79	2	049022	049022 mycoplasma
37	29	46.8	80	5	09TW97	09tw97 plasmodium
38	29	46.8	80	5	09TW76	09tw76 plasmodium
39	29	46.8	80	5	09U004	09u004 plasmodium
40	29	46.8	80	5	09U003	09u003 plasmodium
41	29	46.8	80	5	09U0P5	09u0p5 plasmodium
42	29	46.8	87	10	09FV06	09fv06 phytopthor
43	29	46.8	93	5	018842	018842 caenorhabdi
44	29	46.8	94	10	09CAZ6	09caz6 aradidopsis
45	29	46.8	97	4	09BR02	09br02 homo sapien

## ALIGNMENTS

RESULT 1  
ID 096HT2 PRELIMINARY; PRT; 79 AA.  
AC 096HT2;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE UNKNOWN (PROTEIN FOR MGC:18053).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN, AND GLOBLASTOMA WITH EGFR AMPLIFICATION;  
RA Strausberg R.;  
RL Submitted (Mar-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC008122; AAH08122.1; -; B61779A1A472913B CFC64;  
SQ SEQUENCE 79 AA: 8413 MW; B61779A1A472913B CFC64;

Query Match 59.7%; Score 37; DB 4; Length 79;  
Best Local Similarity 77.8%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 2; Indels 0;

QY 4 PGSHDSQ 12  
Db 64 PGYHDSQ 72

RESULT 2  
ID 037764 PRELIMINARY; PRT; 99 AA.  
AC 037764;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).  
GN COI.  
OS Mytilus edulis (Blue mussel).

OC Mitochondrion.  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92354892; PubMed=1386586;  
 RA Hoffmann R.J., Boore J.L., Brown W.M.;  
 RT "A novel mitochondrial genome organization for the blue mussel,  
 Mytilus edulis".  
 RL Genetics 131:397-412(1992).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRON ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 FERRICTOCHROME C.  
 CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL; M83762; AA31914.1; -;  
 DR EMBL; M83762; AA31913.1; -;  
 DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 1  
 SQ SEQUENCE 99 AA; 11556 MW; 804A0DF951067E40 CRC64;  
 Query Match 56.5%; Score 35; DB 8; Length 99;  
 Best Local Similarity 66.7%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 PEGSQHDSQ 12  
 Db 68 PEGSHVYSQ 76  
 RESULT 3  
 OY0138 PRELIMINARY; PRT; 77 AA.  
 ID 09QYJ8  
 AC 09QYJ8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE DECAF ACCELERATING FACTOR (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WISTARST;  
 RX MEDLINE=20130146; PubMed=10663575;  
 RA Miwa T., Okada N., Okada H.;  
 RT "Alternative exon usage in the 3' region of a single gene generates  
 glycosylphosphatidylinositol-anchored and transmembrane forms of rat  
 decay-accelerating factor".  
 RL Immunogenetics 51:129-137(2000).  
 DR EMBL; AB026905; BAA88994.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 77  
 SQ SEQUENCE 77 AA; 8409 MW; B822DC894E684593 CRC64;  
 Query Match 53.2%; Score 33; DB 11; Length 77;  
 Best Local Similarity 55.6%; Pred. No. 60;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 VEPGSHQHT 9  
 Db 51 VXPATQHV 59  
 RESULT 4  
 OY70B1 PRELIMINARY; PRT; 79 AA.  
 ID 0970B1  
 AC 0970B1;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE IS66 FAMILY ELEMENT, ORF1.  
 GN SP1314.  
 OS Streptococcus pneumoniae.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 pneumoniae".  
 RL Science 293:498-506(2001).  
 DR EMBL; AE007429; AAK75412.1; -;  
 DR TIGR; SP1314; -;  
 KW Complete proteome.  
 SQ SEQUENCE 79 AA; 9260 MW; B5319A03D3235F62 CRC64;  
 Query Match 53.2%; Score 33; DB 16; Length 79;  
 Best Local Similarity 50.0%; Pred. No. 62;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 VEPGSHQHTSQ 12  
 Db 26 VEPGSHRPPDSK 37  
 RESULT 5  
 OY7P24 PRELIMINARY; PRT; 79 AA.  
 ID 097P24  
 AC 097P24;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE IS66 FAMILY ELEMENT, ORF1.  
 GN SP1443.  
 OS Streptococcus pneumoniae.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus



RT pneumonae. Science 293:498-506(2001).  
 RL EMBL: AE007441; AAK75537.1; -  
 DR TIGR: SP1443; -  
 KW Complete proteome.  
 SQ SEQUENCE 79 AA; 9294 MW; 553CFA03DB89F6A CRC64;

Query Match 53.2%; Score 33; DB 16; Length 79;  
 Best Local Similarity 50.0%; Pred. No. 62;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12  
 ||| | : | : | :  
 DB 26 VEIPOSRRFDSK 37

RESULT 6

O9GMAO PRELIMINARY; PRT; 73 AA.  
 AC O9GMAO;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CDS ANTIGEN (FRAGMENT).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;

RT [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98251471; PubMed=9589566;  
 RA Appleyard G.D., Wilkie B.N.;  
 RT "Porcine CDS gene and gene product identified on the basis of inter-  
 RT species conserved cytoplasmic domain sequences."  
 RL Vet. Immunol. Immunopathol. 60:275-283(1998).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Appleyard G.D., Wilkie B.N.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF291867; AAC02103.1; -  
 FT NON\_TER 1 73  
 FT NON\_TER 1 73  
 SQ SEQUENCE 73 AA; 8144 MW; 82DBD6D10F9B58B8 CRC64;

Query Match 51.6%; Score 32; DB 6; Length 73;  
 Best Local Similarity 41.7%; Pred. No. 89;  
 Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12  
 ||| | : | : | :  
 DB 28 VENPASHVDNE 39

RESULT 7

O9U194 PRELIMINARY; PRT; 93 AA.  
 AC O9U194;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHEICAL 10.0 KDA PROTEIN.  
 GN L2903.04.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Hilbert H., Wedler H., Wedler E., Duesterhoeft A., Ivens A.C.,  
 RA Lawson D., Quail M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=FRIEDLIN; PubMed=9477341;  
 RX MEDLINE=98146435; PubMed=9477341;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome."  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL117319; CAB55518.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 93 AA; 9956 MW; 0E47563209669A61 CRC64;

Query Match 51.6%; Score 32; DB 5; Length 93;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12  
 ||| | : | : | :  
 DB 82 VEVPGDHHEKE 93

RESULT 8

O9FOI8 PRELIMINARY; PRT; 95 AA.  
 AC O9FOI8;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).  
 OS Amaranthus quitensis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Amaranthaceae; Amaranthus.  
 OX NCBI\_TaxID=107609;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPON-COPIA-LIKE RETROTRANSPON;  
 RA He Y., Sun M.;  
 RT "Reverse transcriptase sequence evolution in copia-like  
 RT retrotransposons in plants."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF232993; AAC44333.1; -

DR InterPro: IPR003006; IG\_MHC.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 FT NON\_TER 1 95  
 FT NON\_TER 1 95  
 SQ SEQUENCE 95 AA; 11392 MW; 8E931447E1683C5C CRC64;

Query Match 51.6%; Score 32; DB 10; Length 95;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVPGSOH 8  
 ||| | : | : | :  
 DB 26 EVPGKEH 32

RESULT 9

O9I293 PRELIMINARY; PRT; 74 AA.  
 AC O9I293;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE HYPOTHEICAL PROTEIN PA2021.  
 GN PA2021.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gader R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Bradley L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.R.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004628; AAG05409.1; -;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 74 AA; 8217 MW; EDCE3FF490CID520 CRC64;

Query Match 50.0%; Score 31.5; DB 16; Length 74;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 3; Mismatches 1; Indels 3; Gaps 1;

QY 1 VEPGSOHIDSQ 11  
 DB 14 VEIEGSRHAPYDS 27

RESULT 10  
 ID 084884 PRELIMINARY; PRT; 35 AA.  
 AC 084884;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE SIGMA-E FACTOR REGULATORY PROTEIN (FRAGMENT).  
 GN RSEA.  
 OS *Salmonella typhimurium*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Salmonella*.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2;  
 RA "Figueras-Bossel N., Bossel L.;  
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF001386; AAC26063.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 35 AA; 3948 MW; 804CEA18851F0D0 CRC64;

Query Match 50.0%; Score 31; DB 2; Length 35;  
 Best Local Similarity 50.0%; Pred. No. 62;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSQ 12  
 DB 22 VVPGIOTIGTQ 33

RESULT 11  
 ID 096HY2 PRELIMINARY; PRT; 58 AA.  
 AC 096HY2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE UNKNOWN (PROTEIN FOR MGC:151178).  
 OS *Homo sapiens* (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
 RA Strausberg R.;  
 RL EMBL: BC007973; AA07973.1; -;  
 DR EMBL: BC007973; AA07973.1; -;  
 SQ SEQUENCE 58 AA; 6728 MW; 74192B5A8C3DF98C CRC64;

Query Match 50.0%; Score 31; DB 4; Length 58;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSQ 12  
 DB 16 VVPGIAHTDLO 27

RESULT 12  
 ID 062344 PRELIMINARY; PRT; 65 AA.  
 AC 062344;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE R11.4 PROTEIN.  
 GN R11.4.  
 OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdita; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA McMurtry A.A.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode *C. elegans*: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z81577; CAB04648.1; -;  
 SQ SEQUENCE 65 AA; 6997 MW; 8832363CA8D60784 CRC64;

Query Match 50.0%; Score 31; DB 5; Length 65;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSQ 12  
 DB 21 VEMPAQSTQEQ 32

RESULT 13  
 ID 09MAH3 PRELIMINARY; PRT; 65 AA.  
 AC 09MAH3;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PUTATIVE METALLOTHIONEIN-LIKE PROTEIN.  
 GN GRIP24.  
 OS *Vitis vinifera* (Grape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;  
 OC Vitis.  
 OX NCBI\_TaxID=29760;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. SHIRAZ; TISSUE=FRUIT;  
 RX MEDLINE=20177861; PubMed=10712544;  
 RA Davies C., Robinson S.P.;  
 RT "Differential screening indicates a dramatic change in mRNA profiles  
 RT during grape berry ripening. Cloning and characterization of cDNAs  
 RL encoding putative cell wall and stress response proteins.";  
 RL Plant Physiol. 122:803-812(2000).  
 DR EMBL: AJ237990; CAB85630.1; -;  
 SQ SEQUENCE 65 AA; 6777 MW; B5EA7D8D7B9170D4 CRC64;

Query Match 50.0%; Score 31; DB 10; Length 65;  
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VEPGSOH 8  
 Db 39 MEVPAQH 46

## RESULT 14

019274 PRELIMINARY; PRT; 83 AA.  
 AC 019274;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MNC CLASS II ANTIGEN (FRAGMENT).  
 OS Saguinus oedipus (Cotton-top tamarin).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.  
 OX NCBI\_Taxid=9490;  
 RN [1]  
 RP SEQUENCE OF 3-30 FROM N.A.  
 RC STRAIN-B95-8; TISSUE-BLOOD;  
 RX MEDLINE=97252963; PubMed=9098421;  
 RA Bidwell J.L., Lu P., Wang Y., Zhou K., Clay T.M., Bontrop R.E.;  
 RT "DRB, DQA, DOB and DPB nucleotide sequences of Saguinus oedipus B95-8";  
 RL Eur. J. Immunogenet. 21:67-77(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B95-8; TISSUE-BLOOD;  
 RA de Groot N.N.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF027966; AAB84042.1; -  
 DR InterPro: IPR000353; MHC\_II\_beta.  
 DR Pfam: PF00969; MHC\_II\_beta.1.  
 DR ProDom: PD000328; MHC\_II\_beta.1.  
 KM Glycoprotein; MHC II; Transmembrane.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 83 AA; 9737 MW; 465CF7063360C9BE CRC64;  
 Query Match 50.0%; Score 31; DB 7; Length 83;  
 Best Local Similarity 40.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 VPGSOHDSQ 12  
 Db 46 LPAEHLNSQ 55

## RESULT 15

015962 PRELIMINARY; PRT; 87 AA.  
 AC 015962;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE THYROTROPIN BETA SUBUNIT (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93103017; PubMed=8416675;  
 RA Peele M.E., Carr F.E., Baker J.R.Jr., Wartofsky L., Burman K.D.;  
 RT "TSH beta subunit gene expression in human lymphocytes.";  
 RL Am. J. Med. Sci. 305:1-7(1993).  
 DR EMBL: S51112; AAB24571.2; -  
 DR HSSP: P01233; 1XU.  
 DR InterPro: IPR000359; Cys\_knot.  
 DR InterPro: IPR001545; Glyco\_hormone\_beta.  
 DR Pfam: PF00007; Cys\_knot; 1.

DR SMART; SM00068; GHB; 1.  
 DR PROSITE; PS00689; GLYCO\_HORMONE\_BETA\_2; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 87 AA; 9911 MW; 13AEE7F9728D930 CRC64;

Query Match 50.0%; Score 31; DB 4; Length 87;  
 Best Local Similarity 62.5%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEPGSOH 8  
 Db 34 VEIPGCPH 41

Search completed: October 24, 2002, 15:36:51  
 Job time : 15.9672 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:17 : Search time 19.2787 Seconds  
(without alignments)  
69.138 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62  
Sequence: 1 VEVPGSQHTDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	12	21	AAV87461 Cholera toxin B/en
2	62	100.0	15	10	AAV93498 CTP3 epitope of th
3	62	100.0	15	16	AAV85125 Cholera toxin B an
4	62	100.0	21	21	AAV87462 Cholera toxin B/en
5	62	100.0	21	21	AAV87463 E. coli heat labil
6	62	100.0	23	16	AAV87678 Residues 50-64 of
7	62	100.0	26	4	AAV30265 Sequence of amino
8	62	100.0	41	6	AAV50439 Network polymer wh
9	62	100.0	46	6	AAV50436 Network polymer wh
10	62	100.0	47	4	AAV30600 Sequence of amino
11	62	100.0	93	16	AAV72545 ADP-ribosylating t

12	62	100.0	93	20	AAV41816 Escherichia coli v
13	62	100.0	93	20	AAV95226 E. coli heat-labile
14	62	100.0	93	21	AAV68365 Heat labile toxin
15	62	100.0	93	22	AAV66239 E coli verotoxin-1
16	62	100.0	103	6	AAV50340 Sequence of sub-un
17	62	100.0	103	17	AAV04857 Synthetic cholera
18	62	100.0	103	17	AAV94939 Heat labile entero
19	62	100.0	103	17	AAV06606 Cholera toxin B su
20	62	100.0	103	17	AAV06607 Cholera toxin B su
21	62	100.0	103	19	AAV80808 Amino acid sequenc
22	62	100.0	103	22	AAV62365 V. cholera cholera
23	62	100.0	103	22	AAV62367 V. cholera cholera
24	62	100.0	103	22	AAV62370 V. cholera cholera
25	62	100.0	103	22	AAV62379 E. coli LTB protel
26	62	100.0	105	22	AAV62369 V. cholera cholera
27	62	100.0	118	11	AAV04163 Cholera toxin B-su
28	62	100.0	123	22	AAV62374 E. coli LTB protel
29	62	100.0	123	22	AAV62377 E. coli LTB protel
30	62	100.0	124	10	AAV93561 B subunit of the h
31	62	100.0	124	17	AAV06605 Cholera toxin B su
32	62	100.0	124	19	AAV59770 Amino acid sequenc
33	62	100.0	124	21	AAV96652 Plant-optimized E.
34	62	100.0	124	21	AAV96687 Plant-optimized V.
35	62	100.0	124	22	AAV65992 Cholera toxin B su
36	62	100.0	124	22	AAV62359 V. cholera strain
37	62	100.0	124	22	AAV62363 V. cholera cholera
38	62	100.0	124	22	AAV62368 V. cholera cholera
39	62	100.0	124	22	AAV62373 E. coli LTB protel
40	62	100.0	124	22	AAV62375 E. coli LTB protel
41	62	100.0	124	22	AAV62376 E. coli LTB protel
42	62	100.0	124	22	AAV62378 E. coli LTB protel
43	62	100.0	124	22	AAV62380 E. coli LTB protel
44	62	100.0	125	22	AAV62372 E. coli LTB protel
45	62	100.0	126	12	AAV12630 GtfB-1/CTB chimera

## ALIGNMENTS

RESULT 1	AAV87461 standard; peptide: 12 AA.
ID	AAV87461
XX	
AC	AAV87461:
XX	
DT	03-JUL-2000 (first entry)
XX	
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.
XX	
KW	Cholera toxin subunit B: CTxB; heat labile enterotoxin subunit B; EtxB;
KW	beta-4-alpha-2 loop; GW-1 ganglioside receptor; Immunomodulation;
KW	adjuvant; Immune disorder; diarrhoea.
XX	
OS	Vibrio cholerae.
OS	Escherichia coli.
XX	
PN	W0200014114-A1.
XX	
PD	16-MAR-2000.
XX	
PF	07-SEP-1999; 99WC-G802970.
XX	
PK	07-SEP-1998; 98GB-0019484.
XX	
PA	(UYBR-) UNIV BRISTOL.
XX	
PI	Williams NA, Hirst TR:
XX	
DR	WPI; 2000-256943/22.
XX	
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as
PT	immunomodulators and for treating diarrhoea and which do not bind the
PT	glycolipid receptor GM-1 -

XX Disclosure: Page 15; 62pp; English.

CC The invention relates to peptide fragments of the Escherichia coli heat  
CC labile enterotoxin (Etx) and its closely related homologue, cholera  
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
CC composed of one A subunit and five identical B subunits. The A subunit  
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
CC facilitate the entry of subunit A into the host cell via the binding and  
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
CC for some of the effects of Etx and Ctx, it has been found that certain  
CC effects of the toxins, such as immunomodulation, are not mediated  
CC through GM-1 binding. The peptides of the invention are fragments of the  
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
CC Therefore, the peptides may be used in the production of a composition  
CC for treating, preventing and/or modulating a disease associated with an  
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
CC represent preferred peptides of the invention, AAY87460 being  
CC particularly preferred.

XX Sequence 12 AA:

Query Match 100.0%; Score 62; DB 21; Length 12;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSGQHDSQ 12  
| | | | | | | | | | | | | |  
Db 1 VEPGSGQHDSQ 12

RESULT 2  
AAP93498  
ID AAP93498 standard; protein; 15 AA.

XX AAP93498;  
XX  
XX  
XX 03-MAY-1990 (first entry)  
XX  
XX CTP3 epitope of the Cholera toxin B subunit.  
XX  
XX CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;  
XX vaccine; immunotherapy; ds;  
XX  
XX W08910967-A.  
XX  
XX 16-NOV-1989.  
XX  
XX  
XX 05-MAY-1989; 89WO-US01932.  
XX  
XX 05-MAY-1988; 88US-0190570.  
XX  
XX (PRAX-) PRAXIS BIOLOGICS INC.  
XX (STRD ) LEYLAND STANDFORD JUNIOR UNIV.  
XX  
XX Marjarian WR, Stocker BAD, Newton SMC;  
XX  
XX WPI; 1989-356496/48.  
XX N-PSDB; AAN92414.  
XX  
XX New recombinant flagellin gene including sequence - for heterologous  
XX PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.  
XX of antibodies.  
XX  
XX Disclosure; fig. 4B; 137pp; English.  
XX  
XX This sequence corresponds to the CTP3 epitope of the Cholera toxin B  
XX subunit. The DNA sequence encoding this ligates to othersynthetic

CC oligonucleotides to form a new recombinant gene. This encodes  
CC a flagellin fusion protein which can be used in vaccines for immuno-  
CC therapy.

XX Sequence 15 AA:

Query Match 100.0%; Score 62; DB 10; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.9e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSGQHDSQ 12  
| | | | | | | | | | | | | |  
Db 1 VEPGSGQHDSQ 12

RESULT 3  
AAR85125  
ID AAR85125 standard; peptide; 15 AA.

XX AAR85125;  
XX  
XX  
XX 13-JUN-1996 (first entry)  
XX  
XX  
XX  
XX Cholera toxin B antigenic peptide fragment CTP3.  
XX  
XX  
XX Conjugate; cholera; B toxin; peptide fragment; microparticulate;  
XX inert carrier; modified silica; thyroglobulin; oral vaccine;  
XX immunisation; infection; insoluble; digestive tract; antigen;  
XX intestines; antibodies; secretory; Iga class.  
XX  
XX  
XX Vibrio cholerae.  
XX  
XX W09529701-A1.  
XX  
XX  
XX 09-NOV-1995.  
XX  
XX  
XX 02-MAY-1995; 95WO-EP01661.  
XX  
XX 03-MAY-1994; 94IL-0109519.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX  
XX Marks RS, Mirelman D, Sela M;  
XX  
XX WPI; 1995-403805/51.  
XX  
XX  
XX  
XX Vaccines for oral immunisation against infecting agents, e.g.  
XX PT cholera - comprise a conjugate of an antigen of an infecting agent  
XX PT covalently bound to micro-particulate inert carrier, e.g. modified  
XX PT aldehyde silica  
XX  
XX  
XX  
XX Claim 7; Page 25; 40pp; English.  
XX  
XX A compsn. comprising a conjugate of an antigenic cholera B toxin  
XX peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently  
XX bound to a microparticulate inert carrier (e.g. modified silica or  
XX CC thyroglobulin) can be used as an oral vaccine for immunisation  
XX against cholera infection. The inert carrier is insoluble in the  
XX CC digestive tract, allowing presentation of the antigen in the  
XX intestines, where it will elicit antibodies mainly of the  
XX CC secretory Iga class.  
XX  
XX  
XX Sequence 15 AA:

Query Match 100.0%; Score 62; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.9e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSGQHDSQ 12  
| | | | | | | | | | | | | |  
Db 1 VEPGSGQHDSQ 12

```

RESULT 4
AAV87462
ID AAV87462 standard; peptide; 21 AA.
XX
AC AAV87462;
XX
DT 03-JUL-2000 (first entry)
XX
DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.
XX
KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;
KM beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW adjuvant; immune disorder; diarrhoea.
XX
OS Vibrio cholerae.
XX
PN WO200014114-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-GB02970.
XX
PR 07-SEP-1998; 98GB-0019484.
XX
PI (UYBR-) UNIV BRISTOL.
XX
PI Williams NA, Hirst TR;
XX
DR WPI; 2000-256943/22.
XX
PT Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -
XX
PS Disclosure; Page 15; 62pp; English.
XX
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (CtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAV87460-787463
CC represent preferred peptides of the invention, AAV87460 being
CC particularly preferred.
XX
SQ Sequence 21 AA;
XX
XX
Query Match 100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEVPGSQHDSQ 12
DB 6 VEVPGSQHDSQ 17
XX
RESULT 5
AAV87463
ID AAV87463 standard; peptide; 21 AA.

```

```

XX
AC AAV87463;
XX
DT 03-JUL-2000 (first entry)
XX
DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.
XX
KW Heat labile enterotoxin subunit B; EtxB;
KM beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW adjuvant; immune disorder; diarrhoea.
XX
OS Escherichia coli.
XX
PN WO200014114-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-GB02970.
XX
PR 07-SEP-1998; 98GB-0019484.
XX
PI (UYBR-) UNIV BRISTOL.
XX
PI Williams NA, Hirst TR;
XX
DR WPI; 2000-256943/22.
XX
PT Derivatives of Escherichia coli heat labile enterotoxins useful as
PT immunomodulators and for treating diarrhea and which do not bind the
PT glycolipid receptor GM-1 -
XX
PS Disclosure; Page 15; 62pp; English.
XX
CC The invention relates to peptide fragments of the Escherichia coli heat
CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC composed of one A subunit and five identical B subunits. The A subunit
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC ribosyltransferase activity, while the B subunits (CtxB and CtxB)
CC facilitate the entry of subunit A into the host cell via the binding and
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC for some of the effects of Etx and Ctx, it has been found that certain
CC effects of the toxins, such as immunomodulation, are not mediated
CC through GM-1 binding. The peptides of the invention are fragments of the
CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC Therefore, the peptides may be used in the production of a composition
CC for treating, preventing and/or modulating a disease associated with an
CC immune disorder and/or toxin-induced diarrhoea. Sequences AAV87460-787463
CC represent preferred peptides of the invention, AAV87460 being
CC particularly preferred.
XX
SQ Sequence 21 AA;
XX
XX
Query Match 100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VEVPGSQHDSQ 12
DB 6 VEVPGSQHDSQ 17
XX
RESULT 6
AAR76748
ID AAR76748 standard; Protein; 23 AA.
XX
AC AAR76748;
XX
DT 18-MAR-1996 (first entry)

```

```

XX DE Residues 50-64 of cholera toxin B subunit and FimH 224-226.
XX XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX KW FimH; FimH; receptor binding site; PCR; amplify. ss.
XX OS Chimeric - Vibrio cholerae.
XX OS Chimeric - Escherichia coli.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 1 /note= "Represents FimH residue 224"
XX FT Peptide 2..4 /note= "Linker peptide"
XX FT Peptide 5..19 /note= "Cholera toxin B subunit 50-64"
XX FT Peptide 20..22 /note= "Linker peptide"
XX FT Misc-difference 23 /note= "Represents FimH residue 226"
XX FT
XX XX W09520657-A1.
XX PN 03-AUG-1995.
XX PD 27-JAN-1995; 95WO-DK00042.
XX PF 27-JAN-1995; 95WO-DK00042.
XX PR 27-JAN-1994; 94US-0187166.
XX PA (GXBI-) GX BIOSYSTEMS AS.
XX PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX DR WPI; 1995-275442/36.
XX DR N-PSDB; AAQ93061.
XX XX
XX PT Receptor specific bacterial adhesins - useful for targeting active
XX PS compounds and microbial cells to locations of receptors
XX PS
XX PS Example 3; Page 58; 152pp; English.
XX CC This sequence is encoded by a fragment of the the plasmid pHPA93
XX CC which was used in the production of fimH fusion genes comprising
XX CC the cholera toxin B subunit inserted into the fimH gene. This insert
XX CC shows the inclusion of the B subunit into the fimH protein at position
XX CC 224-226. The chimeric genes were then opt. further modified by insertion
XX CC of the hepatitis B virus surface antigen pre-S2 region into a different
XX CC position of the fimH adhesin of type 1 fimbriae. Restriction site handles
XX CC (BglII-sites) were introduced into the fimH gene, and the foreign
XX CC epitopes are then inserted in-frame. In the selected positions the
XX CC function of the epitopes did not significantly alter the adhesive
XX CC on the surface of fimbriae on bacterial hosts illustrated the possibility
XX CC of using bacterial adhesins as general presenters of foreign antigens and
XX CC epitopes. These chimeric genes may be used in the production of variant
XX CC fimH adhesins which may be useful for targeting active compounds
XX CC and microbial cells to locations comprising selected receptors to which
XX CC the adhesins bind.
XX XX
XX SO Sequence 23 AA:
XX
XX Query Match 100.0%; Score 62; DB 16; Length 23;
XX Best Local Similarity 100.0%; Pred. No. 9.6e-05;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VEVPSQHIDSQ 12
XX DB 5 VEVPSQHIDSQ 16
XX
XX RESULT 7
XX AAP30265
XX ID AAP30265 standard; Protein: 26 AA.

```

```

--XX
XX AC AAP30265;
XX XX
XX DT 21-APR-1992 (first entry)
XX XX
XX DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
XX DE carries an Arg at posns. 67 and 73.
XX XX
XX KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
XX KW
XX OS Vibrio cholerae.
XX OS
XX PN EP95426-A.
XX PD 30-NOV-1983.
XX PF 26-MAY-1983; 83EP-0401052.
XX PR 26-MAY-1982; 82FR-0009167.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX PA (INSP ) INST PASTEUR.
XX PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
XX PI Guyongruaz A, Delmas A;
XX DR WPI; 1983-834645/49.
XX XX
XX PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and
XX PT medicaments - effective against Escherichia coli and Vibrio
XX PT cholerae infections, are prep. by solid phase peptide synthesis
XX PS
XX PS Claim 7; Page 11; 13pp; French.
XX CC
XX CC The inventors claim cholera toxin B1 subunit sequences which carry
XX CC Arg in posns. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
XX CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
XX CC these toxins to cell walls. The peptides are used in the treatment of,
XX CC and vaccination against, cholera infections and animal and human
XX CC infections due to E. coli (enterotoxin LT). The medicament may be
XX CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
XX CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
XX CC without carriers. Unit dose when used as a medicament is 50-500mg as
XX CC a vaccine 1-10mg of active cpd.
XX CC
XX SO Sequence 26 AA:
XX
XX Query Match 100.0%; Score 62; DB 4; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.00011;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VEVPSQHIDSQ 12
XX DB 1 VEVPSQHIDSQ 12
XX
XX RESULT 8
XX AAP50439
XX ID AAP50439 standard; protein: 41 AA.
XX AC AAP50439;
XX XX
XX DT 01-JAN-1980 (first entry)
XX XX
XX DE Network polymer which comprises a series of composite E. coli heat-
XX DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
XX XX
XX KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
XX OS Synthetic.
XX XX
XX PN W08502611-A.
XX XX

```



PD 20-JUN-1985.  
 XX 12-DEC-1984; 84MO-US02030.  
 XX 12-DEC-1983; 83US-0559469.  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 PA Houghten RA;  
 PI WPI; 1985-159230/26.  
 DR  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 PS Claim 8; Page 100; 120pp; English.  
 XX The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 41 AA;

Query Match 100.0%; Score 62; DB 6; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 0.00018;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
 |||||||||  
 DB 14 VEVPGSQHIDSQ 25

RESULT 9  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 AC AAP50436;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX  
 KM Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO8502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PF 12-DEC-1984; 84MO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 PS Claim 8; Page 100; 120pp; English.  
 XX The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the

CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 46 AA;

Query Match 100.0%; Score 62; DB 6; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 0.00021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
 |||||||||  
 DB 16 VEVPGSQHIDSQ 27

RESULT 10  
 AAP30600  
 ID AAP30600 standard; protein; 47 AA.  
 AC AAP30600;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 35, 67 and 73.  
 XX  
 KM Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 VIBRIO cholerae.  
 XX  
 EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82FR-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 DR WPI; 1983-834645/49.  
 XX

PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 8; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against, cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX

SQ Sequence 47 AA;

Query Match 100.0%; Score 62; DB 4; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 0.00021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
 |||||||||  
 DB 22 VEVPGSQHIDSQ 33

```

RESULT 11
AAR72545
ID AAR72545 standard; peptide; 93 AA.
XX
AC AAR72545;
XX
DT 28-NOV-1995 (first entry)
XX
DE ADP-ribosylating toxin (verotoxin-1 B-subunit).
XX
KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;
KM active site; E. coli heat labile toxin; verotoxin-1;
KW Bordetella pertussis vaccines.
XX
OS Bacteria sp.
XX
PN EP646599-A.
XX
PD 05-APR-1995.
XX
PF 23-AUG-1994; 94EP-0306219.
XX
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (UNIV ALBERTA.
XX
PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI Oomen R, Read RJ, Stein PE;
XX
DR WPI; 1995-132623/18.
XX
PT New modified forms of pertussis holotoxin - developed using
PT crystalline forms of pertussis holotoxin and its complexes with
PT other molecules
XX
PS Disclosure; Fig 5; 54pp; English.
XX
CC AAR72540-72545 are structurally equivalent B-subunits from three
CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat
CC labile toxin (LT), and verotoxin-1 (VT). The structural
CC information obtd. from these comparisons was used to identify
CC sites which contribute to PT's biological activity. By modifying
CC these sites the claimed PT mutants of the invention were produced,
CC they can be used in the development of vaccines against Bordetella
CC pertussis infection.
XX
SQ Sequence 93 AA;
Query Match 100.0%; Score 62; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 40 VEVPGSQHDSQ 51

```

```

XX
OS Escherichia coli.
XX
PN US5965385-A.
XX
PD 12-OCT-1999.
XX
PF 06-JUN-1995; 95US-0467974.
XX
PR 22-AUG-1994; 94US-0292968.
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (UNIV ALBERTA.
XX
PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
PI Hazes B, Oomen RP;
XX
DR WPI; 1999-579908/49.
XX
PT New method for producing modified pertussis holotoxin -
PT Example 3; Fig 5; 41pp; English.
XX
PS A method has been developed of producing a modified pertussis holotoxin,
XX involving analysis of the 3-dimensional form of the crystalline
XX holotoxin. The pertussis holotoxin modification process comprises:
XX (1) identification of at least one amino acid (aa) residue of the
XX holotoxin for modification by analysing the 3-dimensional form of the
XX crystalline holotoxin, in relation to known information of the protein
XX structure and function; (2) effecting mutagenesis (by removing or
XX replacing a nucleotide sequence encoding at least one (aa) of a tox
XX operon; and (3) expressing mutant tox box in a Bordetella organism to
XX produce the modified holotoxin. This method is used for modifying
XX pertussis holotoxin, by studying its 3-dimensional crystalline
XX structure. Modifying the holotoxin, alters its biological properties.
XX By analysing the 3-dimensional crystalline structure of the pertussis
XX holotoxin, functional (aa) which affect biological properties of the
XX pertussis holotoxin can be identified. This can be used to predict (aa)
XX which contribute to the toxicity of the holotoxin to produce
XX immunoprotective, genetically-detoxified analogues of pertussis
XX holotoxin. The present sequence represents an ADP-ribosylating toxin
XX B-subunit peptide used in the exemplification of the present
XX invention.
XX
SQ Sequence 93 AA;
Query Match 100.0%; Score 62; DB 20; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 40 VEVPGSQHDSQ 51

```

```

RESULT 12
AAV41816
ID AAV41816 standard; peptide; 93 AA.
XX
AC AAV41816;
XX
DT 08-DEC-1999 (first entry)
XX
DE Escherichia coli verotoxin-1 B-subunit.
XX
KW ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;
KW three-dimensional structure; LT; immunoprotective; infection.

```

```

RESULT 13
AAW95226
ID AAW95226 standard; peptide; 93 AA.
XX
AC AAW95226;
XX
DT 16-MAR-1999 (first entry)
XX
DE E. coli heat-labile toxin (LT) beta-subunit sequence.
XX
KW pertussis holotoxin; PT; modified; effector; toxicity; cell binding;
KW enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;
KW structural analysis; interacting site; mitogenicity; adjuvantcity;
KW heat-labile; LT.
XX
OS Escherichia coli.

```

```

XX  US5856122-A.
PN
XX
PD  05-JAN-1999.
XX
PF  22-AUG-1994; 94US-0292968.
XX
PR  22-AUG-1994; 94US-0292968.
PR  24-AUG-1993; 93US-0110947.
PR  31-MAY-1994; 94US-0251121.
XX
PA  (UYAL-) UNIV ALBERTA.
XX
PI  Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI  Oomen RP, Read RJ, Stein PE;
XX
DR  WPI; 1999-105104/09.
XX
PT  Modifications to e.g. enzymatic activity, mitogenicity and cell
PT  binding of pertussis holotoxin - by identifying interaction sites of
PT  a molecule with crystalline toxin and modifying the identified site
XX
PS  Example 3; Fig 5; 40pp; English.
XX
CC  The invention relates to methods of preparing a pertussis holotoxin (PT)
CC  having a modified biological activity. One method comprises identifying
CC  at least 1 site in a PT that interacts with a molecule that is capable of
CC  forming a complex with the holotoxin and which molecule is an effector
CC  molecule which is an adenine nucleotide and which site contributes to
CC  toxicity, cell binding or enzymatic activity of PT. The functional
CC  interacting site(s) are identified by analysing the three dimensional
CC  structure of crystalline PT, determined by X-ray crystallography. The
CC  identified interacting site(s) are modified to alter toxicity, cell
CC  binding or enzyme activity of the PT. The methods can be used to alter a
CC  biological activity such as toxicity, enzymatic activity, mitogenicity,
CC  cell binding and adjuvanticity of the PT. The three-dimensional structure
CC  of PT have functional and/or structural resemblance to other bacterial
CC  toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
CC  heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
CC  sequence represents the beta-subunit of LT toxin.
XX
SQ  Sequence 93 AA;

Query Match 100.0%; Score 62; DB 20; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 VEVPGSQHDSQ 12
DB  40 VEVPGSQHDSQ 51

RESULT 14
AAY68365
ID  AAY68365 standard; Peptide; 93 AA.
XX
AC  AAY68365;
XX
DT  17-APR-2000 (first entry)
XX
DE  Heat labile toxin B subunit SEQ ID NO:26.
XX
KM  Pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;
KM  diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;
KM  infection; crystal structure; X-ray crystallography; detoxification;
KM  immunogenic.
XX
OS  Escherichia coli.
XX
XX  US6018022-A.
PN
XX  25-JAN-2000.
PD
XX

```

```

PF  06-JUN-1995; 95US-0467976.
XX
PR  22-AUG-1994; 94US-0292968.
PR  24-AUG-1993; 93US-0110947.
PR  31-MAY-1994; 94US-0251121.
XX
PA  (CONN-) CONNAUGHT LAB LTD.
XX
PI  Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
PI  Hazes B, Oomen RP;
XX
DR  WPI; 2000-136703/12.
XX
PT  Crystalline form of isolated pertussis holotoxin useful in studying
PT  proteins which have functional resemblance -
XX
PS  Example 3; Fig 5; 42pp; English.
XX
CC  The present invention describes a crystalline form of isolated
CC  pertussis holotoxin, in which the molecules of pertussis toxin have
CC  a three dimensional structure represented in the specification,
CC  complexed with a polysaccharide molecule capable of forming a complex
CC  with the holotoxin. The crystalline form of the pertussis holotoxin
CC  can be used in a comparison with other proteins which have functional
CC  resemblance to pertussis holotoxin with the aim of modifying other
CC  proteins. Identifying the unknown sites of toxicity by comparison
CC  with the three dimensional structure of pertussis holotoxin provides a
CC  technique for detoxification of toxins to produce useful immunogenic
CC  but non-toxic analogues. It can also be used as a primary standard to
CC  measure the quantity, purity or efficacy of less pure compositions
CC  containing pertussis toxin. AAY68340 to AAY68385 represent peptides
CC  used in the exemplification of the present invention.
XX
SQ  Sequence 93 AA;

Query Match 100.0%; Score 62; DB 21; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 VEVPGSQHDSQ 12
DB  40 VEVPGSQHDSQ 51

RESULT 15
AAB66239
ID  AAB66239 standard; Protein; 93 AA.
XX
AC  AAB66239;
XX
DT  03-APR-2001 (first entry)
XX
DE  E coli verotoxin-1 B subunit SEQ ID NO: 26.
XX
KM  Pertussis toxin; crystal structure; whooping cough; biological activity;
KM  lymphocytosis-promoting factor; histamine-sensitising factor;
KM  islet-activating protein.
XX
OS  Escherichia coli.
XX
PN  US6168928-B1.
XX
XX  02-JAN-2001.
PD
XX  21-MAY-1998; 98US-0082514.
XX
PR  22-AUG-1994; 94US-0292968.
PR  24-AUG-1993; 93US-0110947.
PR  31-MAY-1994; 94US-0251121.
XX
XX  (CONN-) CONNAUGHT LAB LTD.
XX

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:18:53 ; Search time 7.08197 Seconds  
(Without alignments)  
41.388 Million cell updates/sec

Title: US-09-786-648-3  
Perfect score: 62  
Sequence: 1 VEPGSGQHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	62	100.0	93	2	US-08-292-968-26
2	62	100.0	93	2	US-08-467-974-26
3	62	100.0	93	2	US-08-467-974-26
4	62	100.0	93	3	US-08-467-976-26
5	62	100.0	93	4	US-08-467-976-26
6	62	100.0	102	3	US-08-952-337-5
7	62	100.0	102	3	US-08-952-337-6
8	62	100.0	103	2	US-08-472-171-2
9	62	100.0	103	2	US-08-894-526-2
10	62	100.0	103	2	US-09-013-047-2
11	62	100.0	103	4	US-09-374-597-2
12	62	100.0	103	4	US-09-191-852-21
13	62	100.0	103	5	PCT-US95-13376-21
14	62	100.0	123	3	US-08-952-337-1
15	62	100.0	123	3	US-08-952-337-2
16	62	100.0	124	2	US-08-747-410-2
17	62	100.0	371	2	US-08-829-026A-6
18	56	90.3	124	1	US-08-449-045C-4
19	56	90.3	124	2	US-08-435-605A-12
20	56	90.3	124	6	5223610-3
21	39	62.9	346	2	US-08-602-359A-34
22	37	59.7	448	2	US-08-878-989-2
23	37	59.7	448	4	US-09-272-796-2
24	37	59.7	508	4	US-09-344-700-2
25	36	58.1	855	2	US-09-027-337-2
26	34	54.8	93	4	US-09-069-023-15
27	34	54.8	219	4	US-09-069-023-12

28	34	54.8	459	6	5194375-6
29	34	54.8	775	2	US-08-714-070A-1
30	33	53.2	321	3	US-09-039-609-4
31	33	53.2	458	3	US-09-039-609-2
32	33	53.2	646	4	US-09-625-188-10
33	33	53.2	845	1	US-08-416-950-11
34	33	53.2	845	2	US-08-469-830-11
35	32	51.6	282	1	US-07-672-304-3
36	32	51.6	297	1	US-08-011-388B-4
37	32	51.6	297	1	US-08-464-051-4
38	32	51.6	297	2	US-08-462-498-4
39	32	51.6	297	3	US-08-554-385-3
40	32	51.6	321	4	US-09-171-461-22
41	32	51.6	360	1	US-08-674-612-3
42	32	51.6	360	1	US-08-469-421-14
43	32	51.6	360	1	US-08-250-975-14
44	32	51.6	360	2	US-08-920-296-3
45	32	51.6	360	2	US-08-605-002A-14

## ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 3656122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COONEY, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & Mcburney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24 AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael T.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Patent No. 5194375  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 2, Appli  
Sequence 10, Appli  
Sequence 11, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 22, Appli  
Sequence 3, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 3, Appli  
Sequence 14, Appli

US-08-292-968-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
|||||  
DB 40 VEVPGSQHIDSQ 51

## RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385

## GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,974

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/467,536

FILING DATE: 06-JUN-1995

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-454 MIS.vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-974-26

DB 40 VEVPGSQHIDSQ 51

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
|||||  
DB 40 VEVPGSQHIDSQ 51

## RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304

## GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,536

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-455 MIS.vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 62; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHIDSQ 12  
|||||  
DB 40 VEVPGSQHIDSQ 51

RESULT 4  
US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022  
; GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,976  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-453 MTS:V9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-976-26

Query Match 100.0%; Score 62; DB 3; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSO 12  
DB 40 VEPGSOHIDSO 51

RESULT 5  
US-09-082-514-26  
Sequence 26, Application US/09082514  
Patent No. 6168928  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,514  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 24-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-810  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-082-514-26

Query Match 100.0%; Score 62; DB 4; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.00037;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSO 12  
DB 40 VEPGSOHIDSO 51

RESULT 6  
US-08-952-337-5  
Sequence 5, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/OD758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-08-952-337-5

Query Match 100.0%; Score 62; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.00036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHIDSO 12  
DB 40 VEPGSOHIDSO 51

Db 49 VEVPSQHTDSQ 60

RESULT 7

US-08-952-337-6

Sequence 6, Application US/08952337

Patent No. 6019973

GENERAL INFORMATION:

APPLICANT: Holmgren, Jan

APPLICANT: Lebens, Michael R.

TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE

FILE REFERENCE: 3846/OD758

CURRENT FILING DATE: 1998-01-05

EARLIER APPLICATION NUMBER: PCT/SE96/00570

EARLIER FILING DATE: 1996-05-02

EARLIER APPLICATION NUMBER: SE 9501682-0

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 102

TYPE: PRT

ORGANISM: Escherichia coli

US-08-952-337-6

Query Match 100.0%; Score 62; DB 3; Length 102;

Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHTDSQ 12

Db 49 VEVPSQHTDSQ 60

RESULT 8

US-08-472-171-2

Sequence 2, Application US/08472171

Patent No. 5932714

GENERAL INFORMATION:

APPLICANT: Loomore, Sheena M.

APPLICANT: Yacoub, Reza K.

APPLICANT: Zealey, Gavin R.

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: Expression Of Gene Products From

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 330 University Avenue, Suite 701

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/472,171

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/393,334

FILING DATE: 23-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: 416-595-1155

TELEFAX: 416-595-1163

TELEX: 065-24567 Simdas

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 103 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-472-171-2

Query Match 100.0%; Score 62; DB 2; Length 103;

Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHTDSQ 12

Db 50 VEVPSQHTDSQ 61

RESULT 9

US-08-894-526-2

Sequence 2, Application US/08894526

Patent No. 5942418

GENERAL INFORMATION:

APPLICANT: Loomore, Sheena M

APPLICANT: Yacoub, Reza K

APPLICANT: Zealey, Gavin R

APPLICANT: Klein, Michel H

TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,526

FILING DATE: 01-DEC-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 103 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-894-526-2

Query Match 100.0%; Score 62; DB 2; Length 103;

Best Local Similarity 100.0%; Pred. No. 0.00036;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHTDSQ 12

Db 50 VEVPSQHTDSQ 61

RESULT 10

US-09-013-047-2



; Sequence 2, Application US/09013047  
; Patent No. 5998168  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/013,047  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/472,171  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334  
; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-013-047-2  
Query Match 100.0%; Score 62; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VEVPGSOHDSQ 12  
Db 50 VEVPGSOHDSQ 61  
RESULT 11  
US-09-374-597-2  
; Sequence 2, Application US/09374597  
; Patent No. 6140082  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, 6th Floor  
; CITY: Toronto

; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/374,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/393,334  
; FILING DATE: FEBRUARY 23, 1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24973  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155  
; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simbas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-374-597-2

Query Match 100.0%; Score 62; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VEVPGSOHDSQ 12  
Db 50 VEVPGSOHDSQ 61  
RESULT 12  
US-09-191-852-21  
; Sequence 21, Application US/09191852  
; Patent No. 6194560  
; GENERAL INFORMATION:  
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq  
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski L.L.P.  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77010  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/191,852  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/13376  
; FILING DATE: 24-OCT-1995  
; APPLICATION NUMBER: 08/817,906  
; FILING DATE: 04-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fox, David L.  
; REGISTRATION NUMBER: 40,612

REFERENCE/DOCKET NUMBER: P015900U51  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-191-852-21

Query Match 100.0%; Score 62; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||  
DB 50 VEVPSQHDSQ 61

RESULT 13  
PCT-US95-13376-21  
Sequence 21, Application PC/US9513376  
GENERAL INFORMATION:  
APPLICANT: The Texas A&M University System  
APPLICANT: 310 Wisedaker  
APPLICANT: College Station, Texas 77843-3369  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones, John W.  
REGISTRATION NUMBER: 31,380  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US95-13376-21

Query Match 100.0%; Score 62; DB 5; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.00036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||  
DB 50 VEVPSQHDSQ 61

RESULT 14  
US-08-952-337-1

Sequence 1, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/OD758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-08-952-337-1

Query Match 100.0%; Score 62; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.00044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||  
DB 70 VEVPSQHDSQ 81

RESULT 15  
US-08-952-337-2  
Sequence 2, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/OD758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-2

Query Match 100.0%; Score 62; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.00044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPSQHDSQ 12  
|||||  
DB 70 VEVPSQHDSQ 81

Search completed: October 24, 2002, 15:23:58  
Job time : 8.08197 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:20:33 : Search time 81.0492 Seconds  
(without alignments)  
52.114 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62  
Sequence: 1 VEVGSGHIDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	12	21	US-09-786-648-3
2	62	100.0	15	11	US-08-732-371-1
3	62	100.0	15	11	US-08-732-371A-1
4	62	100.0	21	21	US-09-786-648-4
5	62	100.0	21	21	US-09-786-648-5
6	62	100.0	93	5	US-08-110-947-10
7	62	100.0	93	5	US-08-110-947A-26

8	62	100.0	93	6	US-08-251-121-26	Sequence 26, Appl
9	62	100.0	103	7	US-08-393-334-2	Sequence 2, Appl
10	62	100.0	103	11	US-08-782-832-15	Sequence 15, Appl
11	62	100.0	103	12	US-08-817-908-21	Sequence 21, Appl
12	62	100.0	103	22	US-09-836-433-14	Sequence 10, Appl
13	62	100.0	116	22	US-09-836-433-20	Sequence 22, Appl
14	62	100.0	119	22	US-09-836-433-22	Sequence 22, Appl
15	62	100.0	124	1	PCT-US99-30747-55	Sequence 55, Appl
16	62	100.0	124	1	PCT-US99-30747-57	Sequence 57, Appl
17	62	100.0	124	18	US-09-470-124-55	Sequence 55, Appl
18	62	100.0	124	18	US-09-470-124-57	Sequence 57, Appl
19	62	100.0	138	13	US-08-914-479A-2	Sequence 2, Appl
20	62	100.0	138	13	US-08-914-479A-2	Sequence 2, Appl
21	62	100.0	313	21	US-09-756-983-15	Sequence 15, Appl
22	62	100.0	351	21	US-09-756-983-18	Sequence 18, Appl
23	62	100.0	354	21	US-09-756-983-22	Sequence 22, Appl
24	62	100.0	371	5	US-08-150-305A-3	Sequence 3, Appl
25	62	100.0	371	11	US-08-784-218-6	Sequence 6, Appl
26	62	100.0	371	12	US-08-829-026-5	Sequence 5, Appl
27	56	90.3	124	21	US-09-760-234-7	Sequence 7, Appl
28	56	90.3	382	1	PCT-US01-08582-3	Sequence 3, Appl
29	56	90.3	382	1	PCT-US01-08582-4	Sequence 4, Appl
30	56	90.3	461	14	US-09-051-315-2	Sequence 2, Appl
31	56	90.3	461	18	US-09-423-493-2	Sequence 2, Appl
32	56	90.3	750	18	US-09-402-100-2	Sequence 2, Appl
33	56	90.3	1338	18	US-09-402-100-4	Sequence 4, Appl
34	44	71.0	484	16	US-09-252-991A-29252	Sequence 29252, A
35	44	67.7	101	15	US-09-134-000-3773	Sequence 3773, Ap
36	42	67.7	631	26	US-60-324-109-19962	Sequence 19962, A
37	42	67.7	2339	21	US-09-733-089-18876	Sequence 18876, A
38	42	67.7	2339	16	US-09-816-660-18876	Sequence 18876, A
39	40	64.5	769	22	US-09-252-991A-17737	Sequence 17737, A
40	39	62.9	7	21	US-09-786-648-2	Sequence 2, Appl
41	39	62.9	91	19	US-09-595-298A-470	Sequence 470, App
42	39	62.9	114	20	US-09-620-394B-4327	Sequence 4327, Ap
43	39	62.9	115	19	US-09-595-298A-468	Sequence 468, App
44	39	62.9	131	26	US-60-171-481-1502	Sequence 1502, Ap
45	39	62.9	318	1	PCT-US99-17130-318	Sequence 318, App

#### ALIGNMENTS

RESULT 1  
US-09-786-648-3  
; Sequence 3, Application US/097866648  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Neil Andrew  
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci  
; FILE REFERENCE: 7438  
; CURRENT APPLICATION NUMBER: US/09/786,648  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: PCT/GB99/02970  
; PRIOR FILING DATE: 1999-09-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: MS DOS  
; SEQ ID NO 3  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: E. coli  
; FEATURE:  
; LOCATION: 50...61  
; OTHER INFORMATION: isolated or synthetic E. coli beta4-alpha2 loop fragment derivabl  
; OTHER INFORMATION: human variant E. coli  
US-09-786-648-3

Query Match 100.0%; Score 62; DB 21; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 12; Conservatively 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 VEPGSHIDSO 12

## RESULT 2

US-08-732-371-1  
Sequence 1, Application US/08732371

GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
APPLICANT: SELA, Michael  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
TITLE OF INVENTION: INFECTING AGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371-1

Query Match 100.0%; Score 62; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12  
Db 1 VEPGSHIDSO 12

RESULT 3  
US-08-732-371A-1  
Sequence 1, Application US/08732371A  
GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
APPLICANT: SELA, Michael  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
TITLE OF INVENTION: INFECTING AGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371A  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371A-1

Query Match 100.0%; Score 62; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12  
Db 1 VEPGSHIDSO 12

RESULT 4  
US-09-786-648-4  
Sequence 4, Application US/09786648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
APPLICANT: Hirst, Timothy Raymond  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccl  
TITLE OF INVENTION: Adjuvants  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/GB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 4  
LENGTH: 21  
TYPE: PPT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 45...65  
OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivabl  
OTHER INFORMATION: human variant E. coli  
US-09-786-648-4

Query Match 100.0%; Score 62; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.00079;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSHIDSO 12  
Db 6 VEPGSHIDSO 17

RESULT 5  
US-09-786-648-5

```

; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: Isolated or synthetic EcxB beta4-alpha2 loop fragment derivable
; OTHER INFORMATION: porcine E. coli
US-09-786-648-5

Query Match          100.0%; Score 62; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSOHDSQ 12
Db      6 VEVPGSOHDSQ 17

RESULT 6
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: OOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; US-08-110-947-10

Query Match          100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSOHDSQ 12
Db      40 VEVPGSOHDSQ 51

RESULT 7
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: OOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947A-26

Query Match          100.0%; Score 62; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSOHDSQ 12
Db      40 VEVPGSOHDSQ 51

RESULT 8
US-08-251-121-26
; Sequence 26, Application US/08251121
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: OOMEN, Raymond P

```

APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/251,121  
FILING DATE: 31-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-335  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-251-121-26

Query Match 100.0%; Score 62; DB 6; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.0044;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VEPGSOHDSQ 12  
Db 40 VEPGSOHDSQ 51

RESULT 9  
US-08-393-334-2  
Sequence 2, Application US/08393334  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,334

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-393-334-2

Query Match 100.0%; Score 62; DB 7; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.005;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VEPGSOHDSQ 12  
Db 50 VEPGSOHDSQ 61

RESULT 10  
US-08-782-832-15  
Sequence 15, Application US/08782832  
GENERAL INFORMATION:  
APPLICANT: Aritzen, Charles J.  
APPLICANT: Mason, Hugh S.  
APPLICANT: Haq, Tariq A.  
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC  
TITLE OF INVENTION: BACTERIAL PROTEIN IN TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,832  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ketelberger, Denise M.  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 36170/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-782-832-15

Query Match 100.0%; Score 62; DB 11; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.005;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

## RESULT 11

US-08-817-906-21  
; Sequence 21, Application US/08817906

## GENERAL INFORMATION:

APPLICANT: Charles J. Arntzen, Hugh S. Mason, John D. Clements,  
APPLICANT: and Tariq A. Haq  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77010

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,906  
FILING DATE: 08/04/97  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-817-906-21

Query Match 100.0%; Score 62; DB 12; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.005;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

## RESULT 12

US-09-836-433-14  
; Sequence 14, Application US/09836433

## GENERAL INFORMATION:

APPLICANT: Yuki, Yoshikazu  
APPLICANT: Udaaka, Shigezo  
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
FILE REFERENCE: Not Assigned  
CURRENT APPLICATION NUMBER: US/09/836,433  
CURRENT FILING DATE: 2001-04-16  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 14  
LENGTH: 103  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-09-836-433-14

Query Match 100.0%; Score 62; DB 22; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

## RESULT 13

US-09-836-433-20  
; Sequence 20, Application US/09836433

## GENERAL INFORMATION:

APPLICANT: Yuki, Yoshikazu  
APPLICANT: Udaaka, Shigezo  
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
FILE REFERENCE: Not Assigned  
CURRENT APPLICATION NUMBER: US/09/836,433  
CURRENT FILING DATE: 2001-04-16  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 20  
LENGTH: 116  
TYPE: PRT  
ORGANISM: synthetic construct  
US-09-836-433-20

Query Match 100.0%; Score 62; DB 22; Length 116;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

## RESULT 14

US-09-836-433-22  
; Sequence 22, Application US/09836433

## GENERAL INFORMATION:

APPLICANT: Yuki, Yoshikazu  
APPLICANT: Udaaka, Shigezo  
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
FILE REFERENCE: Not Assigned  
CURRENT APPLICATION NUMBER: US/09/836,433  
CURRENT FILING DATE: 2001-04-16  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 22  
LENGTH: 119  
TYPE: PRT  
ORGANISM: synthetic construct  
US-09-836-433-22

Query Match 100.0%; Score 62; DB 22; Length 119;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
| | | | | | | | | |  
DB 50 VEVPGSQHDSQ 61

## RESULT 15

PCT-US99-30747-55  
; Sequence 55, Application PC/TUS9930747

APPLICANT: Boyce Thompson Institute for Plant Research at Cor  
TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In  
FILE REFERENCE: 4868/85427  
CURRENT APPLICATION NUMBER: PCT/US99/30747

;  
; CURRENT FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 124  
; TYPE: PRF  
; ORGANISM: Escherichia coli  
PCT-US99-30747-55

Query Match 100.0%; Score 62; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 0.0062;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSOHDSQ 12  
| | | | | | | | | |  
Db 71 VEVPGSOHDSQ 82

Search completed: October 24, 2002, 15:30:56  
Job time : 81.0492 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:21:08 ; Search time 27.1475 Seconds  
(without alignments)  
132.044 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEPGSGHDSQ 12

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 1105779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents, AA, New: \*  
1: /cgn2\_6/ptodata/2/paa/US06\_NEM\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEM\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEM\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEM\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEM\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEM\_COMB.pep: \*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEM\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	103	US-09-791-537-19387	Sequence 19387, A
2	62	100.0	103	US-09-791-537-38611	Sequence 38611, A
3	62	100.0	103	US-09-791-537-38639	Sequence 38639, A
4	62	100.0	103	US-09-791-537-68591	Sequence 68591, A
5	62	100.0	103	US-09-791-537-74385	Sequence 74385, A
6	62	100.0	103	US-10-110-364-8	Sequence 8, Appl
7	62	100.0	103	US-10-110-364-10	Sequence 10, Appl
8	62	100.0	103	US-10-110-364-13	Sequence 13, Appl
9	62	100.0	103	US-10-110-364-22	Sequence 22, Appl
10	62	100.0	104	US-09-791-537-87980	Sequence 87980, A
11	62	100.0	104	US-09-791-537-99772	Sequence 99772, A
12	62	100.0	105	US-10-110-364-12	Sequence 12, Appl
13	62	100.0	113	US-09-791-537-73608	Sequence 73608, A
14	62	100.0	123	US-10-110-364-17	Sequence 17, Appl
15	62	100.0	123	US-10-110-364-20	Sequence 20, Appl
16	62	100.0	124	US-09-791-537-28360	Sequence 28360, A
17	62	100.0	124	US-09-791-537-29489	Sequence 29489, A
18	62	100.0	124	US-09-791-537-33623	Sequence 33623, A
19	62	100.0	124	US-09-791-537-40552	Sequence 40552, A
20	62	100.0	124	US-09-791-537-60743	Sequence 60743, A
21	62	100.0	124	US-09-791-537-78640	Sequence 78640, A
22	62	100.0	124	US-09-791-537-92185	Sequence 92185, A
23	62	100.0	124	US-09-791-537-94644	Sequence 94644, A
24	62	100.0	124	US-09-791-537-103241	Sequence 103241, A
25	62	100.0	124	US-09-791-537-103660	Sequence 103660, A
26	62	100.0	124	US-10-110-364-2	Sequence 2, Appl

27	62	100.0	124	US-10-110-364-6	Sequence 6, Appl
28	62	100.0	124	US-10-110-364-11	Sequence 11, Appl
29	62	100.0	124	US-10-110-364-16	Sequence 16, Appl
30	62	100.0	124	US-10-110-364-18	Sequence 18, Appl
31	62	100.0	124	US-10-110-364-19	Sequence 19, Appl
32	62	100.0	124	US-10-110-364-21	Sequence 21, Appl
33	62	100.0	124	US-10-110-364-23	Sequence 23, Appl
34	62	100.0	125	US-10-110-364-15	Sequence 15, Appl
35	62	100.0	129	US-09-791-537-11854	Sequence 131854, A
36	62	100.0	131	US-09-791-537-130348	Sequence 130348, A
37	62	100.0	138	US-10-141-627-2	Sequence 2, Appl
38	56	90.3	103	US-09-791-537-42610	Sequence 42610, A
39	56	90.3	103	US-09-791-537-129309	Sequence 129309, A
40	56	90.3	103	US-10-110-364-5	Sequence 5, Appl
41	56	90.3	103	US-10-110-364-7	Sequence 7, Appl
42	56	90.3	124	US-09-791-537-123948	Sequence 123948, A
43	56	90.3	124	US-10-110-364-4	Sequence 4, Appl
44	56	90.3	124	US-10-110-364-9	Sequence 9, Appl
45	56	90.3	382	PCR-US02-20978-1	Sequence 1, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-791-537-19387
; Sequence 19387, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Dancer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19387
; LENGTH: 103
; TYPE: PRF
; ORGANISM: pdb 1CTLD
US-09-791-537-19387

Query Match          100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEPGSGHDSQ 12
DB      50 VEPGSGHDSQ 61

RESULT 2
US-09-791-537-38611
; Sequence 38611, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Dancer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38611
; LENGTH: 103
; TYPE: PRF
; ORGANISM: pdb 1CHPD
US-09-791-537-38611
```

```
Query Match          100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSQHDSQ 12
      |||
Db      50 VEVPGSQHDSQ 61

RESULT 3
US-09-791-537-38639
; Sequence 38639, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38639
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1CHOD
US-09-791-537-38639

Query Match          100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSQHDSQ 12
      |||
Db      50 VEVPGSQHDSQ 61

RESULT 4
US-09-791-537-68591
; Sequence 68591, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 68591
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-09-791-537-68591

Query Match          100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSQHDSQ 12
      |||
Db      50 VEVPGSQHDSQ 61

RESULT 5
US-09-791-537-74385
; Sequence 74385, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek

Query Match          100.0%; Score 62; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSQHDSQ 12
      |||
Db      50 VEVPGSQHDSQ 61

RESULT 6
US-10-110-364-8
; Sequence 8, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)-(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 998409
US-10-110-364-8

Query Match          100.0%; Score 62; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00086;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VEVPGSQHDSQ 12
      |||
Db      50 VEVPGSQHDSQ 61

RESULT 7
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
```

```
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 103
; TYPE: PRF
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511
US-10-110-364-10

Query Match
Best Local Similarity 100.0%; Score 62; DB 6; Length 103;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
Db 50 VEVPGSQHDSQ 61

RESULT 8
US-10-110-364-13
; Sequence 13, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 103
; TYPE: PRF
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
US-10-110-364-13

Query Match
Best Local Similarity 100.0%; Score 62; DB 6; Length 103;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
Db 50 VEVPGSQHDSQ 61

RESULT 9
US-10-110-364-22
; Sequence 22, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
```

```
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 103
; TYPE: PRF
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.
US-10-110-364-22

Query Match
Best Local Similarity 100.0%; Score 62; DB 6; Length 103;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
Db 50 VEVPGSQHDSQ 61

RESULT 10
US-09-791-537-87980
; Sequence 87980, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 87980
; LENGTH: 104
; TYPE: PRF
; ORGANISM: pdb 2CHBD
US-09-791-537-87980

Query Match
Best Local Similarity 100.0%; Score 62; DB 5; Length 104;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEVPGSQHDSQ 12
Db 51 VEVPGSQHDSQ 62

RESULT 11
US-09-791-537-99772
; Sequence 99772, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 99772
; LENGTH: 104
; TYPE: PRF
; ORGANISM: pdb 3CHBD
```

US-09-791-537-99772

Query Match 100.0%; Score 62; DB 5; Length 104;  
Best Local Similarity 100.0%; Pred. No. 0.00087;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
|||||  
DB 51 VEPGSOHDSQ 62

RESULT 12

US-10-110-364-12  
; Sequence 12, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT APPLICATION NUMBER: US/10/110,364  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 105  
; TYPE: PRT  
; ORGANISM: Vibrio cholera  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(105)  
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (Ogawa  
US-10-110-364-12

Query Match 100.0%; Score 62; DB 6; Length 105;  
Best Local Similarity 100.0%; Pred. No. 0.00088;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
|||||  
DB 51 VEPGSOHDSQ 62

RESULT 13

US-09-791-537-73608  
; Sequence 73608, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Blonoxix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 73608  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: pdb 1LTRD  
US-09-791-537-73608

Query Match 100.0%; Score 62; DB 5; Length 113;  
Best Local Similarity 100.0%; Pred. No. 0.00096;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
|||||  
DB 50 VEPGSOHDSQ 61

RESULT 14

US-10-110-364-17  
; Sequence 17, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT APPLICATION NUMBER: US/10/110,364  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(123)  
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.  
US-10-110-364-17

Query Match 100.0%; Score 62; DB 6; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPGSOHDSQ 12  
|||||  
DB 71 VEPGSOHDSQ 82

RESULT 15

US-10-110-364-20  
; Sequence 20, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT APPLICATION NUMBER: US/10/110,364  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(123)  
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.  
US-10-110-364-20

Query Match 100.0%; Score 62; DB 6; Length 123;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	VEVPGSOHDSO	12
Db	70	VEVPGSOHDSO	81

Search completed: October 24, 2002, 15:33:19  
Job time : 27.1475 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:17:58 ; Search time 9.2459 Seconds  
(without alignments)  
124.712 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEVPGSQHDSQ 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	62	100.0	124	1	XVVCB
2	62	100.0	124	1	QLECB
3	62	100.0	124	1	QLECB
4	62	100.0	124	1	QLECB
5	62	100.0	124	1	QLECB
6	62	100.0	124	1	QLECB
7	62	100.0	124	1	QLECB
8	62	100.0	124	1	QLECB
9	62	100.0	124	1	QLECB
10	62	100.0	124	1	QLECB
11	62	100.0	124	1	QLECB
12	62	100.0	124	1	QLECB
13	62	100.0	124	1	QLECB
14	62	100.0	124	1	QLECB
15	62	100.0	124	1	QLECB
16	62	100.0	124	1	QLECB
17	62	100.0	124	1	QLECB
18	62	100.0	124	1	QLECB
19	62	100.0	124	1	QLECB
20	62	100.0	124	1	QLECB
21	62	100.0	124	1	QLECB
22	62	100.0	124	1	QLECB
23	62	100.0	124	1	QLECB
24	62	100.0	124	1	QLECB
25	62	100.0	124	1	QLECB
26	62	100.0	124	1	QLECB
27	62	100.0	124	1	QLECB
28	62	100.0	124	1	QLECB
29	62	100.0	124	1	QLECB

30	35	56.5	377	2	T29750	hypothetical prote
31	35	56.5	389	2	A98303	hypothetical prote
32	35	56.5	389	2	AE2880	monooxygenase (limp
33	35	56.5	427	2	UC5694	stress-activated p
34	35	56.5	443	2	T29029	hypothetical prote
35	35	56.5	472	2	AG0432	glutamate synthase
36	35	56.5	539	2	S53529	monophenol monooxy
37	35	56.5	593	2	T24379	hypothetical prote
38	35	56.5	670	2	T02092	hypothetical prote
39	35	56.5	960	2	A41638	beta-fructofuranos
40	35	56.5	961	2	A55380	chitin synthase (E
41	35	56.5	1258	2	T29041	faciogenital dyspl
42	35	56.5	2206	2	G71611	hypothetical prote
43	34	54.8	168	2	T27563	hypothetical prote
44	34	54.8	181	2	T45990	hypothetical prote
45	34	54.8	244	1	T01091	NADH dehydrogenase

#### ALIGNMENTS

RESULT 1  
XVVCB  
Cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N  
N:Alternate names: enterotoxin beta chain  
C:Species: Vibrio cholerae  
C:Date: 24-Apr-1984 #sequence, revision 01-Sep-2000 #text, change 02-Feb-2001  
C:Accession: S14624; S39238; S39241; H82196; JCI078; SI7666; PC1010; A05130; A01819;  
R:Dams, E.; de Wolf, M.; Dierick, W.  
Submitted to the EMBL Data Library, March 1991  
A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol  
A:Reference number: S14623  
A:Accession: S14624  
A:Molecule type: DNA  
A:Residues: 1-124 <DM>  
A:Cross-references: EMBL:X58786; NID:g48420; PIDN:CAA41593.1; PID:g48422  
A:Experimental source: strain 2125  
R:Rebens, M.; Holmgren, J.  
Submitted to the EMBL Data Library, November 1993  
A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera  
A:Reference number: S39238  
A:Accession: S39238  
A:Molecule type: DNA  
A:Residues: 1-124 <LEB>  
A:Cross-references: EMBL:X76390; NID:g433856; PIDN:CAA53973.1; PID:g433857  
A:Accession: S39241  
A:Molecule type: DNA  
A:Residues: 1-124 <LEW>  
A:Cross-references: EMBL:X76391; NID:g433859; PIDN:CAA53976.1; PID:g433861  
R:Rebens, M.; Holmgren, J.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: AB2035; M01D:20406833  
A:Accession: H82196  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-124 <HEI>  
A:Cross-references: GB:AE004224; GB:AE003852; NID:g9655952; PIDN:AAF94613.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.  
Chinese Biotech. J. 9, 395-399, 1993  
A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.ch  
A:Reference number: JCI078  
A:Accession: JCI078  
A:Molecule type: DNA  
A:Residues: 1-20, 'Q', 22-31, 'Q', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>  
A:Experimental source: classical biotype strain 569B  
R:Dams, E.; de Wolf, M.; Dierick, W.  
Biochim. Biophys. Acta 1090, 139-141, 1991  
A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic

A:Accession: S17666  
 A:Molecule type: DNA  
 A:Residues: 1-38, 'H', 40-67, 'T', 69-124 <DA2>  
 A:Cross-references: EMBL:X5785; NID:948888; PIDD:CAA41591.1; PID:g4890  
 R:Ma, O.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
 Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
 A:Title: B subunit of cholera toxin produced in *Escherichia coli*.  
 A:Reference number: PC1010  
 A:Accession: PC1010  
 A:Molecule type: Protein  
 A:Residues: 22-38, 'H', 40-41 <MAQ>  
 R:McKalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.  
 Nature 306, 551-557, 1983  
 A:Reference number: A93320; MUID:84068199  
 A:Accession: A05130  
 A:Molecule type: DNA  
 A:Residues: 1-32, 'S', 34-74, 'S', 76-124 <MEK>  
 A:Cross-references: GB:X00171; NID:948347; PIDD:CAA2496.1; PID:g758351  
 R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
 J. Biol. Chem. 252, 7257-7264, 1977  
 A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
 A:Reference number: A01819; MUID:78005537  
 A:Accession: A01819  
 A:Molecule type: Protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <KUR>  
 R:Li, C.Y.  
 J. Biol. Chem. 252, 7249-7256, 1977  
 A:Title: Determination of the primary structure of cholera toxin B subunit.  
 A:Reference number: A38033; MUID:78005536  
 A:Accession: A38033  
 A:Molecule type: Protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69, 'E', 71-90, 'N', 92-124 <LAI>  
 A:Note: The difference at residue 70 may be due to deamidation during preparation  
 R:Nakashima, Y.; Napierkowski, P.; Schaefer, D.E.; Konigsberg, W.H.  
 FEBS Lett. 68, 275-278, 1976  
 A:Title: Primary structure of the B subunit of cholera enterotoxin.  
 A:Reference number: A38034; MUID:77026365  
 A:Accession: A38034  
 A:Molecule type: Protein  
 A:Residues: 22-38, 'H', 40-67, 'T', 69, 'E', 71, 'Q', 74-75, 'V', 78-86, 'Q', 88-99, 'Q', 101-103, 'Q'  
 R:Takao, T.; Watanabe, H.; Shimonishi, Y.  
 Eur. J. Biochem. 146, 503-508, 1985  
 A:Title: Facile identification of protein sequences by mass spectrometry.  
 A:Reference number: A21910; MUID:85126976  
 A:Accession: A21910  
 A:Molecule type: Protein  
 A:Residues: 22-38, 'H', 40-42, 'N', 44-67, 'T', 69-90, 'N', 92-124 <TAK>  
 A:Experimental source: biotype Inaba 569b  
 A:Note: Asn-65 was partially deaminated to Asp  
 C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
 C:Genetics:  
 A:Gene: VC1456  
 A:Map position: 1  
 C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and  
 ciate noncovalently with the subunit B, an aggregate of five beta chains  
 C:Function:  
 A:Description: Involved in binding of the toxin to cell membranes  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin; toxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
 DB 71 VEVPGSQHDSQ 82  
 RESULT 2

OUECB  
 heat-labile enterotoxin chain B precursor - *Escherichia coli*  
 C:Species: *Escherichia coli*  
 C:Date: 29-Jun-1981 #sequence revision 29-Jun-1981 #text change 18-Jun-1999  
 C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
 R:Dallas, W.S.; Palkov, S.  
 Nature 288, 499-501, 1980  
 A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli* heat  
 A:Reference number: A01820; MUID:81074965  
 A:Accession: A01820  
 A:Molecule type: mRNA  
 A:Residues: 1-124 <DLR>  
 R:Ivanov, T.; Gojovic, T.; Yokota, T.  
 J. Bacteriol. 169, 1352-1357, 1987  
 A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia coli*  
 A:Reference number: A26946; MUID:87137303  
 A:Accession: B26946  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>  
 A:Cross-references: EMBL:M5363; NID:9148335; PIDD:AAA24792.1; PID:g148336  
 R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
 Infect. Immun. 48, 73-77, 1985  
 A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
 A:Reference number: I41194; MUID:85156481  
 A:Accession: I41194  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122  
 A:Cross-references: GB:M17874; NID:9145830; PIDD:AAA98064.1; PID:g145831  
 R:Experimental source: plasmid ENT-R PCG86  
 R:Ibrahim, I.; Gentz, R.  
 J. Biol. Chem. 262, 10189-10194, 1987  
 A:Title: A functional interaction between the signal peptide and the translation appa  
 lliculum.  
 A:Reference number: I41287; MUID:87280041  
 A:Accession: I41287  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-22 <RE2>  
 A:Cross-references: GB:M17101; NID:9146375; PIDD:AAA23973.1; PID:g146376  
 R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
 FEBS Microbiol. Lett. 108, 157-161, 1993  
 A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
 A:Reference number: I3542; MUID:93252225  
 A:Accession: I67644  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R  
 A:Cross-references: GB:S60731; NID:9408994; PIDD:AA60441.1; PID:g408996  
 R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
 Microb. Pathog. 2, 381-390, 1987  
 A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin 1  
 A:Reference number: A61475; MUID:89180953  
 A:Accession: A61475  
 A:Molecule type: Protein  
 A:Residues: 22-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-95, 'A', 97-122, 'E', 124 <TSU>  
 A:Experimental source: strain 240-3  
 C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six  
 C:Function:  
 A:Description: the biological activity of the toxin is produced by the A chain, which  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEVPGSQHDSQ 12  
 DB 71 VEVPGSQHDSQ 82



## RESULT 3

hypothetical protein SC2A11.21c SC2A11.21c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T34767

R:Murphy, L.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z21556

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-242 <MUR>

A:Cross-references: EMBL:AL031184; PIDN:CAA20190.1; GSPDB:GN00070; SCODEB:SC2A11.21c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCODEB:SC2A11.21c

Query Match 64.5%; Score 40; DB 2; Length 242;

Best Local Similarity 70.0%; Pred. No. 6.1;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10  
|||||:|

Db 204 VEPGSDHID 213

## RESULT 4

protein F20B17.2 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse ear cross)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C96827

R:Theologis, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

Nansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, J.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; M0ID:21016719

A:Accession: C96827

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-260 <STO>

A:Cross-references: GB:AE005173; NID:g7715588; PIDN:AAF68106.1; GSPDB:GN00141

C:Genetics:

A:Gene: F20B17.2

A:Map position: 1

Query Match 62.9%; Score 39; DB 2; Length 260;

Best Local Similarity 50.0%; Pred. No. 10;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 VEPGSOHID 12  
|||||:|

Db 65 VEPGSOHID 76

## RESULT 5

T14755

hypothetical protein DKFZP564A0122.1 - human

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14755

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wilmann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14755

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <WAM>

A:Cross-references: EMBL:AL110209

A:Experimental source: fetal brain; clone DKFZP564A0122

C:Genetics:

A:Note: DKFZP564A0122.1

Query Match

Best Local Similarity 62.9%; Score 39; DB 2; Length 272;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 VEPGSOHID 10  
|||||:|

Db 246 ELPSEHIE 254

## RESULT 6

H75446

(S)-2-hydroxy-acid oxidase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000

C:Accession: H75446

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; M0ID:20036896

A:Accession: H75446

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-353 <WHI>

A:Cross-references: GB:AE001954; GB:AE000513; NID:g6458751; PIDN:AAF10604.1; PID:g645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1031

A:Map position: 1

C:Superfamily: (S)-2-hydroxy-acid oxidase; (S)-2-hydroxy-acid oxidase homology

F3-297/Domain: (S)-2-hydroxy-acid oxidase homology <2H>

Query Match 62.9%; Score 39; DB 2; Length 353;

Best Local Similarity 75.0%; Pred. No. 14;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 VPGSOHID 10  
|||||:|

Db 189 VPGSEHID 196

## RESULT 7

A35781

hippocampus-derived neurotrophic factor precursor - rat

N:Alternate names: neurotrophin-3 precursor

C:Species: Rattus norvegicus (Norway rat)

C:Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 16-Jul-1999

C:Accession: A35781; A40094

R:Ernfors, P.; Ibanez, C.F.; Ebendal, T.; Olson, L.; Persson, H.

Proc. Natl. Acad. Sci. U.S.A. 87, 5454-5458, 1990

A:Title: Molecular cloning and neurotrophic activities of a protein with structural s

A:Reference number: A35781; M0ID:90319130

A:Accession: A35781

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-282 <ERN>

A:Cross-references: GB:M34643

R:Maizompiere, P.C.; Belluscio, L.; Squinto, S.; Ip, N.Y.; Furth, M.E.; Lindsay, R.M.

Science 247, 1446-1451, 1990

A:Title: Neurotrophin-3: a neurotrophic factor related to NGF and BDNF.

A:Reference number: A40094; M0ID:90208301

A:Accession: A40094

A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 25-282 <MAT>  
 A:Cross-references: GB:M33968; NID:g205771; PIDN:AAA41727.1; PID:g205772  
 C:Superfamily: nerve growth factor beta chain

Query Match 61.3%; Score 38; DB 2; Length 282;  
 Best Local Similarity 54.5%; Pred. No. 17;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPGSQHDS 11  
 Db 1 VDVPGNSHTDA 11

RESULT 8  
 A41670

carbon-monoxide dehydrogenase (EC 1.2.99.2) beta chain [similarity] - Clostridium therm  
 C:Species: Clostridium thermaceticum  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000  
 C:Accession: A41670

R:Korton, T.A.; Runquist, J.A.; Ragsdale, S.W.; Shanmugasundaram, T.; Wood, H.G.; Ljung  
 J. Biol. Chem. 266, 23824-23828, 1991

A:Title: The primary structure of the subunits of carbon monoxide dehydrogenase/acetyl-C  
 A:Reference number: A41670; MUID:92084676

A:Accession: A41670  
 A>Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-674 <MOR>  
 A:Cross-references: GB:M62727; NID:g144784; PIDN:AAA23228.1; PID:g144785

C:Superfamily: carbon-monoxide dehydrogenase beta chain; hybrid cluster (4Fe-2S-30) hom  
 C:Keywords: 4Fe-2S-30; 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein; c  
 F:256-606/Domain: hybrid cluster (4Fe-2S-30) homology <HCL>

F:66,71,76,90/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
 F:263,321,355,470,500,550,585/Binding site: Ni-3Fe-2S-30 cluster (His, Glu, Cys, Cys, Cy

F:470/Modified site: cysteine persulfide (Cys) #status predicted

Query Match 61.3%; Score 38; DB 1; Length 674;  
 Best Local Similarity 54.5%; Pred. No. 45;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVPGSQHDSQ 12  
 Db 379 KIPGAVHIDQ 389

RESULT 9  
 B85518

hypothetical protein 20347 [imported] - Escherichia coli (strain O157:H7, substrain EDL5  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: B85518  
 R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: B85518  
 A>Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-352 <STO>

A:Cross-references: GB:AE005174; NID:g12513064; PIDN:AA654606.1; GSPDB:GN00145; UWGP:203  
 A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:  
 A:Gene: 20347

Query Match 59.7%; Score 37; DB 2; Length 352;  
 Best Local Similarity 60.0%; Pred. No. 35;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPGSQHID 10  
 Db 320 VVVGANHND 329

RESULT 10  
 T01751

gibberellin 20-oxidase - common tobacco  
 N:Alternate names: Ntc16 protein

C:Species: Nicotiana tabacum (common tobacco)  
 C>Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 20-Jun-2000

C:Accession: T01751  
 R:Tanaka-DeGuchi, M.; Itoh, H.; Oyama, N.; Koshioke, M.; Matsuo, M.

submitted to the EMBL Data Library, July 1998  
 A:Description: Over-expression of a tobacco homeobox gene, NTH15, decreases the expre

A:Reference number: Z14418  
 A:Accession: T01751

A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA

A:Residues: 1-367 <TRAN>  
 A:Cross-references: EMBL:AB016084

C:Genetics:  
 A:Gene: Ntc16  
 C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 59.7%; Score 37; DB 2; Length 367;  
 Best Local Similarity 87.5%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVPGSQHI 9  
 Db 149 EVPGSQHI 156

RESULT 11  
 S17246

chromate synthase (EC 4.6.1.4) - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein G2501; protein YGL148w

C:Species: Saccharomyces cerevisiae  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: S17246; S64162  
 R:Jones, D.G.U.; Reusser, U.; Braus, G.H.

Mol. Microbiol. 5, 2143-2152, 1991  
 A:Title: Molecular cloning, characterization and analysis of the regulation of the AR

A:Reference number: S17246; MUID:92114793  
 A:Accession: S17246

A:Molecule type: DNA  
 A:Residues: 1-376 <CON>

A:Cross-references: EMBL:X60190; NID:g3386; PIDN:CAA42745.1; PID:g3387  
 R:Volckaert, G.; Voet, M.; Verhasselt, P.; Defoor, E.

submitted to the Protein Sequence Database, May 1996  
 A:Reference number: S64153

A:Accession: S64162  
 A:Molecule type: DNA

A:Residues: 1-376 <VOL>  
 A:Cross-references: EMBL:Z72670; NID:g1322731; PIDN:CAA96860.1; PID:g1322732; GSPDB:G

A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:ARO2; MIPS:YGL148w

A:Cross-references: SGD:S0003116; MIPS:YGL148w  
 A:Map position: 7L

C:Superfamily: chorismate synthase  
 C:Keywords: phosphorus-oxygen lyase; transmembrane protein

F:347-363/Domain: transmembrane #status predicted <TM>

Query Match 59.7%; Score 37; DB 1; Length 376;  
 Best Local Similarity 70.0%; Pred. No. 37;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPGSQHID 10  
 Db 264 VVPGSKHND 273

RESULT 12  
 F90667  
 hypothetical protein EC90310 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: F90667  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shingawa, H. DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: F90667  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-378 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA833733.1; PID:q13359767; GSPDB:GN00154  
 A:Experimental source: strain 0157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECs0310

Query Match 59.7%; Score 37; DB 2; Length 378;  
 Best Local Similarity 60.0%; Pred. No. 37;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10  
 | | | | | | | |  
 Db 346 VVPGANHVD 355

RESULT 13  
 G96554  
 hypothetical protein F19C24.16 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G96554  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizart, L. Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis. A:Reference number: A66141; MUID:21016719  
 A:Accession: G96554  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-423 <STO>  
 A:Cross-references: GB:AB005173; NID:q11094753; PIDN:AMG29686.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F19C24.16  
 A:Map position: 1

Query Match 59.7%; Score 37; DB 2; Length 423;  
 Best Local Similarity 75.0%; Pred. No. 42;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 VPGSOHID 10  
 | | | | | | | |  
 Db 403 VPGIOHVD 410

RESULT 14  
 T39683  
 zootin-like protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
 C:Accession: T39683; T40195  
 R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, March 1998  
 A:Reference number: Z21869  
 A:Accession: T39683  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-124 <OLI>  
 A:Cross-references: EMBL:AL049489; PIDN:CAB39796.1; GSPDB:GN00067; SPDB:SPBC1778.01C  
 A:Experimental source: strain 972h-; cosmid cl778  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hlibert, H.; Duesterhoeft, A submitted to the EMBL Data Library, February 1998  
 A:Reference number: T40195  
 A:Accession: T40195  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 89-442 <WOO>  
 A:Cross-references: EMBL:Z97992; PIDN:CAB10796.1; GSPDB:GN00067; SPDB:SPBC30D10.01  
 A:Experimental source: strain 972h-; cosmid G30D10  
 C:Genetics:  
 A:Gene: SPDB:SPBC1778.01C; SPDB:SPBC30D10.01  
 A:Map position: 2

Query Match 59.7%; Score 37; DB 2; Length 442;  
 Best Local Similarity 50.0%; Pred. No. 44;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 EYVPGSOHIDS 11  
 | | | | | | | |  
 Db 371 DVPSAEHVD 380

RESULT 15  
 T01770  
 hypothetical protein A\_IG002P16.9 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T01770  
 R:Miller, N.; Beck, C.; Kramer, J. submitted to the EMBL Data Library, June 1997  
 A:Description: The sequence of A. thaliana IG002P16. A:Reference number: Z14421  
 A:Accession: T01770  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-550 <MLL>  
 A:Cross-references: EMBL:AF007270; NID:q2191157; PID:q2191172; GSPDB:GN00063; ATSP:A\_IG002P16.9  
 A:Gene: ATSP:A\_IG002P16.9  
 A:Map position: 5  
 A:Introns: 159/1; 272/1; 434/2; 477/3

Query Match 59.7%; Score 37; DB 2; Length 550;  
 Best Local Similarity 66.7%; Pred. No. 56;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 VEPGSOHIDSQ 12  
 | | | | | | | |  
 Db 260 VTCSGSOHIDFQ 271

Search completed: October 24, 2002, 15:23:19  
 Job time : 12.2459 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:42 : Search time 4.72131 Seconds  
(without alignments)  
98.412 Million cell updates/sec

Title: US-09-786-648-3

Perfect score: 62

Sequence: 1 VEVPGSQHSDS 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	124	1	CHTB_VIBCH
2	62	100.0	124	1	ELBH_ECOLI
3	62	100.0	124	1	ELBP_ECOLI
4	38	61.3	674	1	DCMB_MOORH
5	37	59.7	376	1	AROC_YEAST
6	36	58.1	240	1	YBSA_THEMA
7	36	58.1	455	1	GDDH_BACSU
8	36	58.1	743	1	BGAL_THRET
9	36	58.1	855	1	ST14_HUMAN
10	36	58.1	1785	1	GIS3_YEAST
11	35	56.5	302	1	PHEB_PSESP
12	35	56.5	313	1	YF68_CAEEL
13	35	56.5	371	1	YMA8_PSEAE
14	35	56.5	539	1	TYRO_ASPOB
15	35	56.5	670	1	INVL_MAIZE
16	35	56.5	960	1	CHS3_NEUCR
17	35	56.5	960	1	FED1_MOUSE
18	35	56.5	961	1	FSD1_HUMAN
19	34	54.8	219	1	CIDA_HUMAN
20	34	54.8	244	1	NUHM_ARATH
21	34	54.8	365	1	FTZ3_PYRAB
22	34	54.8	365	1	FTZ3_PYRHO
23	34	54.8	419	1	GSC_DROME
24	34	54.8	432	1	AROC_NEUCR
25	34	54.8	456	1	SR54_THEAC
26	34	54.8	459	1	IL7R_MOUSE
27	34	54.8	491	1	CD5_RAT
28	34	54.8	500	1	GABT_HUMAN
29	34	54.8	508	1	Y202_HUMAN
30	34	54.8	557	1	HLXB_SERMA
31	34	54.8	560	1	INR1_SHEEP
32	34	54.8	775	1	TH1L_SCHPO
33	34	54.8	842	1	DPOL_HPBVA

34	34	54.8	1151	1	Y245_TREPA
35	33.5	54.0	375	1	MASP_MOUSE
36	33.5	54.0	375	1	MASP_MOUSE
37	33	53.2	259	1	MOB2_YEAST
38	33	53.2	264	1	FTSQ_STRCU
39	33	53.2	291	1	CORC_BUCAT
40	33	53.2	386	1	YMAN_BACSU
41	33	53.2	454	1	VNIC_THOGV
42	33	53.2	504	1	A37C_DROME
43	33	53.2	534	1	YOG1_CAEEL
44	33	53.2	621	1	HEM1_AGABI
45	33	53.2	666	1	PDI4_MOUSE

## ALIGNMENTS

RESULT 1	ID	CHTB_VIBCH	STANDARD	PRT	124 AA.
AC	P01556	O9J002	(Rel. 01, Created)		
DT	21-JUL-1986				
DT	13-AUG-1987		(Rel. 05, Last sequence update)		
DT	16-OCT-2001		(Rel. 40, Last annotation update)		
DE	Cholera enterotoxin, beta chain precursor.				
GN	CYRX OR TOXB OR VCI456.				
OS	Vibrio cholerae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.				
OX	NCBI_TaxID=666;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=84061784; PubMed=6315707;				
RT	Lockman H., Kaper J.B.;				
RT	"Nucleotide sequence analysis of the A2 and B subunits of Vibrio cholerae enterotoxin.";				
RL	J. Biol. Chem. 258:13722-13726(1983).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EL TOR 2125;				
RC	MEDLINE=84068199; PubMed=6646234;				
RA	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,				
RT	de Wilde M.;				
RT	"Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development.";				
RL	Nature 306:551-557(1983).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=4260B / SEROTYPE O139;				
RC	MEDLINE=94237453; PubMed=8181723;				
RA	Lebens M., Holmgren J.;				
RT	"Structure and arrangement of the cholera toxin genes in Vibrio cholerae O139.";				
RL	FEMS Microbiol. Lett. 117:197-202(1994).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=1854 / O139-BENGAL;				
RC	Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,				
RA	Honda T.;				
RT	Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.				
RN	[6]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;				
RC	MEDLINE=20406833; PubMed=10952301;				
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,				
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,				
RA	Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,				
RA	Ernolava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,				
RA	McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,				

O83273	treponema p
P70124	mus musculu
P70564	rattus norv
P43563	saccharomyc
O86038	streptomyce
P57518	buchnera ap
P37535	bacillus su
P89216	thogoto vir
P16487	dtrosophila
P34610	caenorhabd
O92403	agaricus bi
O92183	mus musculu

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 RN [7]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005537; PubMed=903363;  
 RA Kirovsky A., Markel D.E., Peterson J.W.;  
 RT Covalent structure of the beta chain of cholera enterotoxin.";  
 RL J. Biol. Chem. 252:7257-7264(1977).  
 RN [8]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005536; PubMed=903362;  
 RA Lai C.-Y.;  
 RT "Determination of the primary structure of cholera toxin B subunit.";  
 RL J. Biol. Chem. 252:7249-7256(1977).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=94272319; PubMed=8003954;  
 RA Merritt E.A., Sarfaly S., van den Akker F., I'Hoir C., Martial J.A.,  
 RA Hol W.G.J.;  
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
 RT pentasaccharide.";  
 RL Protein Sci. 3:166-175(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=95387394; PubMed=7658472;  
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,  
 RA Maulik P.R., Reed R.A., Shipley G.G.;  
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:  
 RT choleraenoid.";  
 RL J. Mol. Biol. 251:550-562(1995).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX STRAIN-OGAMA 41 / CLASSICAL BIOTYPE;  
 RX MEDLINE=97376625; PubMed=9232653;  
 RA Merritt E.A., Sarfaly S., Jobling M.G., Chang T., Holmes R.K.,  
 RA Hirst T.R., Hol W.G.J.;  
 RT "Structural studies of receptor binding by cholera toxin mutants.";  
 RL Protein Sci. 6:1516-1528(1997).  
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
 CC BINDING TO CELL MEMBRANES.  
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS, AN ALPHA AND A GAMMA CHAIN  
 CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
 CC DISULFIDE BOND. ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
 CC 6 BETA CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 DR EMBL; X00171; CAA24996.1; -;  
 DR EMBL; K01170; AAA27573.1; -;  
 DR EMBL; D30053; BAA06291.1; -;  
 DR EMBL; X58786; CAA41593.1; -;  
 DR EMBL; X76390; CAA53973.1; -;  
 DR EMBL; X76391; CAA53976.1; -;  
 DR EMBL; AE004224; AAF94613.1; -;  
 DR PIR; A01819; XYVCB.  
 DR PIR; A05130; A05130.  
 DR PIR; S14624; S14624.  
 DR PDB; 2CHB; 03-DEC-97.  
 DR PDB; 3CHB; 12-AUG-98.  
 DR PDB; 1CHP; 08-MAR-96.  
 DR PDB; 1CHQ; 08-MAR-96.  
 DR PDB; 1FGB; 23-DEC-96.  
 DR PDB; 1XPB; 01-APR-97.  
 DR PDB; 1XRC; 01-AUG-96.

DR PDB; 1CT1; 15-OCT-97.  
 DR TIGR; VCI456; -;  
 DR InterPro; IPR001835; Enterotoxin\_B.  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PRO0772; ENTEROTOXIN.  
 DR ProDom; PD012805; Enterotoxin\_B; 1.  
 KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124  
 FT DISULFID 30 107  
 FT CONFLICT 33 33  
 FT CONFLICT 39 39  
 FT CONFLICT 43 43  
 FT CONFLICT 68 68  
 FT CONFLICT 70 70  
 FT CONFLICT 75 75  
 FT CONFLICT 91 91  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT TURN 34 35  
 FT STRAND 36 44  
 FT STRAND 47 51  
 FT TURN 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT HELIX 80 99  
 FT TURN 100 100  
 FT STRAND 102 109  
 FT STRAND 115 123  
 SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EABE3EBF CRC64;  
 Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 6e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 VEPFGSQHDSQ 12  
 Db 71 VEPFGSQHDSQ 82  
 ID ELBH\_ECOLI STANDARD; PRT; 124 AA.  
 AC P13811;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).  
 GN ELTB OR LTPB.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-ISOLATE H74-114;  
 RC MEDLINE=85156481; PubMed=3884513;  
 RX Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT cistrons from Escherichia coli of human and porcine origin.";  
 RL Infect. Immun. 48:73-77(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H10407;  
 RX MEDLINE=83114628; PubMed=6759877;  
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
 RT "Overlapping genes in the heat-labile enterotoxin operon originating  
 RT from Escherichia coli human strain.";  
 RL Mol. Gen. Genet. 188:356-359(1982).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN-ISOLATE H10407;  
 RX MEDLINE=93252225; PubMed=8486242;  
 RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;  
 RT "Amino acid sequence of heat-labile enterotoxin from chicken  
 RT enterotoxigenic Escherichia coli is identical to that of human strain  
 RT H 10407.";  
 RL FEMS Microbiol. Lett. 108:157-161(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ETEC LT 87;  
 RA Germani Y., Desperrier J.M.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=95349400; PubMed=7623669;  
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
 RX MEDLINE=99185101; PubMed=10085117;  
 RA Matkovic-Galogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,  
 RA Tossi A., Palu G., Zanotti G.;  
 RT "Crystal structure of the B subunit of Escherichia coli heat-labile  
 RT enterotoxin carrying peptides with anti-herpes simplex virus type 1  
 RT activity.";  
 RL J. Biol. Chem. 274:8764-8769(1999).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M17874; AAA98064.1; -;  
 DR EMBL; J01646; AAB02982.1; -;  
 DR EMBL; S60731; AAB60441.1; -;  
 DR EMBL; X83966; CAA58800.1; -;  
 DR PDB; 1LTR; 23-MAR-99.  
 DR InterPro; IPR001835; Enterotoxin\_B.  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR ProDom; PD012805; Enterotoxin\_B; 1.  
 DR Enterotoxin; Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT VARIANT 6 6 F -> C (IN ISOLATE H10407).  
 FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).  
 FT VARIANT 34 34 H -> R (IN ISOLATE H10407).  
 SQ SEQUENCE 124 AA; 14027 MW; E9EF7C7B9D3BC47 CRC64;

Query Match 100.0%; Score 62; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. NO. 6e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEPYSGSHDSQ 12  
 DB 71 VEPYSGSHDSQ 82

RESULT 3  
 ELBP\_ECOLI STANDARD; PRT; 124 AA.  
 AC P32890; P13768; P01557;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1995 (Rel. 32, last sequence update)  
 DT 15-JUL-1998 (Rel. 36, last annotation update)  
 DE Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LT-B).  
 GN ELTB OR LTBP.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE P307;  
 RX MEDLINE=81074965; PubMed=7003397;  
 RA Dallas W.S., Falkow S.;  
 RT "Amino acid sequence homology between cholera toxin and Escherichia  
 RT coli heat-labile toxin.";  
 RL Nature 288:499-501(1980).  
 RN [2]  
 RP REVISIONS TO 28 AND 64.  
 RC STRAIN-ISOLATE P307;  
 RX MEDLINE=85156481; PubMed=3884513;  
 RA Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT cistrons from Escherichia coli of human and porcine origin.";  
 RL Infect. Immun. 48:73-77(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE PCG86;  
 RX MEDLINE=87137303; PubMed=3546273;  
 RA Yamamoto T., Gojohori T., Yokota T.;  
 RT "Evolutionary origin of pathogenic determinants in enterotoxigenic  
 RT Escherichia coli and Vibrio cholerae O1.";  
 RL J. Bacteriol. 169:1352-1357(1987).  
 RN [4]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RX MEDLINE=87280041; PubMed=3301830;  
 RA Ibrahim I., Gentz R.;  
 RT "A functional interaction between the signal peptide and the  
 RT translation apparatus is detected by the use of a single point  
 RT mutation which blocks translocation across mammalian endoplasmic  
 RT reticulum.";  
 RL J. Biol. Chem. 262:10189-10194(1987).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE=93240541; PubMed=8478941;  
 RA Sima T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;  
 RT "Refined structure of Escherichia coli heat-labile enterotoxin, a  
 RT close relative of cholera toxin.";  
 RL J. Mol. Biol. 230:890-918(1993).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=91238966; PubMed=2034287;  
 RA Sima T.K., Pronk S.E., Kalk K.H., Martina E.S., van Zanten B.A.M.,  
 RA Witbold B., Hol W.G.J.;  
 RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin  
 RT from E. coli.";  
 RL Nature 351:371-377(1991).  
 RN [7]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=95349400; PubMed=7623669;  
 RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; X60190; CAA42745.1; -;  
 DR EMBL; X99960; CAA68214.1; -;  
 DR EMBL; Z72670; CAA96860.1; -;  
 DR PIR; S17246; S17246.  
 DR SCD; S0003116; ARO2.  
 DR InterPro: IPR000453; Chorismate\_synt.  
 DR Pfam; PF01264; Chorismate\_synt; 1.  
 DR Pfam; PD002941; Chorismate\_synt; 1.  
 DR PROSITE; PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
 DR PROSITE; PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
 DR PROSITE; PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
 KW Lysase; Aromatic amino acid biosynthesis.  
 SO SEQUENCE 376 AA; 40838 MW; AF3AF6505B91EBE CRC64;

Query Match 59.7%; Score 37; DB 1; Length 376;  
 Best Local Similarity 70.0%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHID 10  
 | | | | | | | | | |  
 Db 264 VSPGSKHND 273

RESULT 6  
 YB5A\_THEME STANDARD; PRT; 240 AA.  
 ID YB5A\_THEME  
 AC P58009;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein TM1158.1.  
 GN TM1158.1.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MS8 / DSM 3109;  
 RC MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser J.C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima";  
 RL Nature 399:323-329(1999).  
 RN [2]  
 RP IDENTIFICATION.  
 RA Medigue C., Bocs S.;  
 RL Unpublished observations (APR-2001).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AE001773; -; NOT\_ANNOTATED\_CDS.  
 DR TIGR; TM1158.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 240 AA; 27773 MW; IBEF66C1C8BD2700 CRC64;

Query Match 58.1%; Score 36; DB 1; Length 240;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 VPGSOHID 12  
 | | | | | | | | | |  
 Db 226 VPGSEHLECK 235

RESULT 7  
 GUDH\_BACSU STANDARD; PRT; 455 AA.  
 ID GUDH\_BACSU  
 AC P42238;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable glucarate dehydratase (EC 4.2.1.40) (GDH) (glucD).  
 GN GUDH.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RC MEDLINE=95219079; PubMed=7704254;  
 RX Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;  
 RA "Determination of a 21548 bp nucleotide sequence around the 24  
 RT degrees region of the Bacillus subtilis chromosome.";  
 RL Microbiology 141:269-275(1995).  
 CC -1- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-  
 CC DEOXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-glucarate = 5-dehydro-4-deoxy-D-glucarate +  
 CC H(2)O.  
 CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.  
 CC -1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE  
 CC LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; D30808; BAA06470.1; -;  
 DR EMBL; Z99105; CAB12043.1; -;  
 DR HSSP; P42206; 1B0G.  
 DR Subtilist; BG1161; gudD.  
 DR InterPro: IPR001354; MR\_MLE.  
 DR Pfam; PF01188; MR\_MLE; 1.  
 DR Pfam; PF02746; MR\_MLE\_N; 1.  
 KW Lysase; Complete proteome.  
 SO SEQUENCE 455 AA; 50782 MW; 323846007698C2A CRC64;

Query Match 58.1%; Score 36; DB 1; Length 455;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 EYVPSOHI 9  
 | | | | | | | | | |  
 Db 64 EYVGSBHI 71

RESULT 8  
 BGAL\_THERT STANDARD; PRT; 743 AA.  
 ID BGAL\_THERT  
 AC P77989;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Beta-galactosidase (EC 3.2.1.23) (Lactase).  
 GN LACZ OR LACA.  
 OS Thermoaerobacter ethanolicus (Clostridium thermohydrophilicum).  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Thermoaerobacter group; Thermoaerobacter.  
 OX NCBI\_TaxID=1757;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33223 / 39E;  
 RA Zverlov V.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
 CC galactose residues in beta-D-galactosides.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y08557; CAA69850.1; -  
 DR InterPro: IPR001649; Glyco\_hydro\_2.  
 DR Pfam: PF00703; Glyco\_hydro\_2\_1.  
 DR Pfam: PF02836; Glyco\_hydro\_2\_C; 1.  
 DR Pfam: PF02837; Glyco\_hydro\_2\_N; 1.  
 DR PROSITE: PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
 DR PROSITE: PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
 KM Hydrolyase: Glycosidase.  
 FT ACT\_SITE 388 388 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 453 453 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 743 AA; 85796 MW; FE011F51751DPC CRC64;  
 OY 2 EVPSQHIDSQ 12  
 Db 346 EIPGMHIDE 356  
 Query Match 58.1%; Score 36; DB 1; Length 743;  
 Best local similarity 54.5%; Pred. No. 44;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Yamaguchi N., Mitsui S.;  
 RT "Molecular cloning of a novel transmembrane serine protease expressed  
 RT in human prostate.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,  
 RA O'Brien T.J.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 327-855 FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 340-664 FROM N.A.  
 RA Cao J., Fan W., Zheng S.;  
 RT "Genomic analysis of a novel human serine protease SMC19.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP CHARACTERIZATION.  
 RC TISSUE=Milk;  
 RX MEDLINE=99303582; Pubmed-10373425;  
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
 RT "Purification and characterization of a complex containing matrilysin  
 RT and a Kunitz-type serine protease inhibitor from human milk.";  
 RL J. Biol. Chem. 274:18237-18242(1999).  
 CC -1- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
 CC OR LYS AS THE P1 SITE.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
 CC -1- SIMILARITY: CONTAINS 4 IDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF118224; MAD42765.2; -  
 DR EMBL: AF133086; AAF00109.1; -  
 DR EMBL: AB030036; BAB20376.1; -  
 DR EMBL: AF057145; AAG15395.1; -  
 DR EMBL: BC005826; AAH05826.1; -  
 DR EMBL: AF283256; AAG13949.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.302; -  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR002172; IDL\_recept\_A.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00057; Idl\_recept\_a; 4.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00261; IDLRECEPTOR.  
 DR SMART: SM00042; CUB; 2.  
 DR SMART: SM00192; IDLa; 3.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01209; IDLRA\_1; 2.  
 DR PROSITE: PS00668; IDLRA\_2; 4.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.

```

DR PROSITE: PS00134; TRYPsin HIS; 1.
DR PROSITE: PS00135; TRYPsin SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KW Transmembrane; Repeat.
FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 214 334 COB 1.
FT DOMAIN 340 447 COB 2.
FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 615 854 SERINE PROTEASE.
FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 327 329 FEA -> GTR (IN REF. 5).
FT CONFLICT 381 381 R -> S (IN REF. 4).
FT CONFLICT 674 674 A -> V (IN REF. 3).
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 58.1%; Score 36; DB 1; Length 855;
Best Local Similarity 55.6%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEPGSGHIT 9
DB 370 LEVPMNHV 378

RESULT 10
GLS3_YEAST STANDARD; PRT; 1785 AA.
AC Q04952;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative 1,3-beta-glucan synthase component (EC 2.4.1.34) (1,3-beta-D-
glucan-UDP glucosyltransferase).
GN YMR306W OR YMR952.08.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Connor R., Churcher C.M., Barrall B.G., Raeburn M.A., Walsh S.V.;
RL Submitted (Apr-1995) to the EMBL/Genbank/DDBJ databases.
CC -1- ANALYTIC ACTIVITY: UDP-glucose + [(1,3)-beta-D-glucosyl](N) = UDP
+ [(1,3)-beta-D-glucosyl](N+1).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO GLS1 AND GLS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49212; CA89139.1; -
DR SGD; S0004923; YMR306W.
DR InterPro; IPR003440; Glucan_synthase.
DR Pfam; PF02364; Glucan_synthase; 1.
KW Hypothetical protein; Transmembrane; Transferase; Glycosyltransferase.

```

```

FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 415 435 POTENTIAL.
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 508 528 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 572 592 POTENTIAL.
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 1215 1235 POTENTIAL.
FT TRANSMEM 1268 1288 POTENTIAL.
FT TRANSMEM 1303 1323 POTENTIAL.
FT TRANSMEM 1370 1390 POTENTIAL.
FT TRANSMEM 1394 1414 POTENTIAL.
FT TRANSMEM 1475 1495 POTENTIAL.
FT TRANSMEM 1514 1534 POTENTIAL.
FT TRANSMEM 1549 1569 POTENTIAL.
FT TRANSMEM 1585 1605 POTENTIAL.
FT TRANSMEM 1655 1675 POTENTIAL.
FT TRANSMEM 1713 1733 POTENTIAL.
SQ SEQUENCE 1785 AA; 207482 MW; 3475446DA46C6120 CRC64;

Query Match 58.1%; Score 36; DB 1; Length 1785;
Best Local Similarity 54.5%; Pred. No. 1,2e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EYPGSGHDSQ 12
DB 403 EWPGAGHLSR 413

RESULT 11
PHEB_PSESP STANDARD; PRT; 302 AA.
ID PHEB_PSESP
AC P31019;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Catechol 1,2-dioxygenase (EC 1.13.11.1).
GN PHEB.
OS Pseudomonas sp. (strain EST1001).
OC Plasmid pEST1412.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91192610; PubMed=2013408;
RA Kivisaar M., Kasak L., Nurk A.;
RT *Sequence of the plasmid-encoded catechol 1,2-dioxygenase-expressing
RT gene, pheB, of phenol-degrading Pseudomonas sp. strain EST1001.,"
RL Gene 98:15-20(1991).
CC -1- ANALYTIC ACTIVITY: Catechol + O(2) = cis,cis-muconate.
CC -1- COFACTOR: FERRIC ION.
CC -1- PATHWAY: PHENOL DEGRADATION.
CC -1- SIMILARITY: BELONGS TO THE INTRADIOL RING-CLEAVAGE DIOXYGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57500; AAC64900.1; -
DR PIR; JN0143; JN0143.
DR HSSP; P00437; 3PCD.
DR InterPro; IPR000627; Dioxygenase.
DR Pfam; PF00775; Dioxygenase; 1.
DR PROSITE; PS00083; INTRADIOL_DIOXYGENAS; 1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Dioxygenase; Iron;
KW Plasmid.
FT METAL 164 164 IRON (BY SIMILARITY).

```

FT METAL 198 198 IRON (BY SIMILARITY).  
 FT METAL 222 222 IRON (BY SIMILARITY).  
 FT METAL 224 224 IRON (BY SIMILARITY).  
 SQ SEQUENCE 302 AA; 33362 MW; A86F17E68D1EAC3A CRC64;

Query Match 56.5%; Score 35; DB 1; Length 302;  
 Best Local Similarity 41.7%; Pred. No. 25;  
 Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEPGSGHIDSQ 12  
 Db 227 ISAGHGHITQ 238

RESULT 12  
 ID YP68\_CAEEL STANDARD; PRT; 313 AA.  
 AC 009217;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.  
 OS B0495.8.  
 GN Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Kirsten J.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: SOME, TO YEAST YDL087C AND S.POMME SPC16A11.13.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U21317; AAA62527.1; -  
 DR WormPep: B0495.8; CE01766.  
 DR Hypothetical protein.  
 SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BFED0 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 313;  
 Best Local Similarity 45.5%; Pred. No. 26;  
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 EYPGSGHIDSQ 12  
 Db 11 QLMGSGHVDNK 21

RESULT 13  
 ID YM18\_PSEAE STANDARD; PRT; 371 AA.  
 AC 001609; 0911P9;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein PA2218.  
 GN PA2218.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01 / H103;  
 RX MEDLINE=93051258; PubMed=1427017;

RA Huang H., Stenmel R.J., Bellido F., Rawling E., Hancock R.E.W.;  
 RT "Analysis of two gene regions involved in the expression of the  
 RT impenem-specific, outer membrane porin protein OprD of Pseudomonas  
 RT aeruginosa.";  
 RL FEMS Microbiol. Lett. 76:267-274(1992).

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.V., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).

CC -1- SIMILARITY: TO E.COLI YCJY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z14064; CAAT8447.1; -  
 DR EMBL: AE004648; MAG5606.1; ALT\_INIT.  
 DR PIR: S23861; S23861.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 16 R -> P (IN REF. 1).  
 FT CONFLICT 73 R -> P (IN REF. 1).  
 FT CONFLICT 261 A -> G (IN REF. 1).  
 SQ SEQUENCE 371 AA; 40840 MW; D7EB0CCAC95A7CF6 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 371;  
 Best Local Similarity 62.5%; Pred. No. 32;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSGHID 10  
 Db 343 VPGASHVD 350

RESULT 14  
 ID TYRO\_ASPOR STANDARD; PRT; 539 AA.  
 AC 000234;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Tyrosinase (EC 1.14.18.1) (Monophenol monooxygenase).  
 GN MELO.  
 OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 22788 / RIB 128;  
 RX MEDLINE=95200965; PubMed=7893753;  
 RA Fujita Y., Uraga Y., Ichishima E.;  
 RT "Molecular cloning and nucleotide sequence of the tyrosinase gene,  
 RT melo, from Aspergillus oryzae and expression of the gene in yeast  
 RT cells.";  
 RL Biochim. Biophys. Acta 1261:151-154(1995).  
 CC -1- FUNCTION: THIS IS A COPPER-CONTAINING OXIDASE THAT FUNCTIONS IN  
 CC THE FORMATION OF PIGMENTS SUCH AS MELANINS AND OTHER POLYPHENOLIC  
 CC COMPOUNDS.  
 CC -1- CATALYTIC ACTIVITY: L-tyrosine + L-DOPA + O(2) -> L-DOPA +  
 CC DOPAQUNONE + H(2)O.

```

CC - COFACTOR: BINDS TWO COPPER IONS.
CC - ENZYME REGULATION: ACTIVATED BY ACIDIFYING TREATMENT AT PH 3.0.
CC - SUBUNIT: HOMOTETRAMER.
CC - PM: THE N-TERMINAL IS BLOCKED.
CC - SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D37929; BAA07149.1; -.
DR InterPro: IPR002227; Tyrosinase.
DR Pfam: PF00264; Tyrosinase; 2.
DR PRINTS: PR00092; TYROSINASE.1; 1.
DR PROSITE: PS00497; TYROSINASE.1; 1.
DR PROSITE: PS00498; TYROSINASE.2; 1.
KW Melanin biosynthesis: Oxidoreductase; Monooxygenase; Copper.
FT METAL 63 63 COPPER A (BY SIMILARITY).
FT METAL 84 84 COPPER A (BY SIMILARITY).
FT METAL 93 93 COPPER A (BY SIMILARITY).
FT METAL 290 290 COPPER B (BY SIMILARITY).
FT METAL 294 294 COPPER B (BY SIMILARITY).
FT METAL 333 333 COPPER B (BY SIMILARITY).
FT THIOETH 82 84 BY SIMILARITY.
SQ SEQUENCE 539 AA; 60604 MW; CD2ECD702A018E15 CRC64;

Query Match
Best Local Similarity 56.5%; Score 35; DB 1; Length 539;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 EYVPSQHIDSQ 12
| | | : | | |
DB 360 EYVPSQHIDSQ 370

RESULT 15
INV1_MAIZE STANDARD: PRU; 670 AA.
AC P49175;
DR 01-FEB-1996 (Rel. 33, Created)
DR 01-FEB-1996 (Rel. 33, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-fructofuranosidase 1 precursor (EC 3.2.1.26) (Sucrose-6-phosphate
DE hydrolase 1) (Invertase 1).
GN IVRL.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_ B73;
RX MEDLINE=95357417; PubMed=7630946;
RA Xu J., Pemberton G.H., Almira E.C., McCarty D.R., Koch K.E.;
RT "The ivr 1 gene for invertase in maize.";
RL Plant Physiol. 108:1293-1294(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing beta-D-
CC fructofuranoside residues in beta-D-fructofuranosides.
CC -1- SUBCELLULAR LOCATION: Vacuolar.
CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: U16123; AAA83439.1; -.
DR MaizEDB: 86037; -.
DR InterPro: IPR001362; Glyco_hydro_32.
DR Pfam: PF00251; Glyco_hydro_32; 1.
DR PROSITE: PS00609; GLYCOSYL_HYDROL_F32; 1.
KW Hydrolase; Glycosidase; Glycoprotein; zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 1 POTENTIAL.
FT PROPEP 2 2 POTENTIAL.
FT CHAIN 3 670 BETA-FRUCTOFURANOSIDASE 1.
FT ACT_SITE 139 139 BY SIMILARITY.
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 670 AA; 71932 MW; DED0989C7E6AEB0 CRC64;

Query Match
Best Local Similarity 56.5%; Score 35; DB 1; Length 670;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 VEVPSSQHIDSQ 12
| | | | | | |
DB 61 VVVLASQHDGQ 72

Search completed: October 24, 2002, 15:21:01
Job time : 6.72131 secs

```

**THIS PAGE BLANK (USPTO)**

DT 01-NOV-1996 (TREMBLREl. 01, Created)

```
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL STRAIN 569B;
RA MEDLINE=91355224; PubMed=1883840;
RT Dams E., De Wolf M., Dierick W.;
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae
RT classical strain 569B.";
RL Biochim. Biophys. Acta 1090:139-141(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Shi C., Cao C., Zhang J., Ma O.;
RL Chin. Biochem. J. 9:395-399(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Xu L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X58785; CA441591.1; -.
DR EMBL: U23679; AAC34728.1; -.
DR EMBL: A00931; CA000098.1; -.
DR HSSP: P01556; ZCHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 124 AA; 13919 MW; D6BF83FF7924EA3 CRC64;
Query Match 100.0%; Score 62; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 71 VEVPGSQHDSQ 82
RESULT 3
Q9RP15 PRELIMINARY; PRT; 124 AA.
AC Q9RP15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA ENTEROTOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNIH002;
RA Shin H.-J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RT cassette from Vibrio cholerae KNIH002 isolated in Korea.";
RL Misalimurag Holji 35:205-210(1999).
DR EMBL: AF175708; AAD51360.1; -.
DR HSSP: P01556; ZCHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FF793E5B9 CRC64;
```

```
Query Match 100.0%; Score 62; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 71 VEVPGSQHDSQ 82
RESULT 4
Q56635 PRELIMINARY; PRT; 124 AA.
AC Q56635;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA TOXIN PRECURSOR.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.,
RA Honda T.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae
RT strains 854 (O139-bengal) and S7 (O37) from two outbreaks.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: D30052; BAA06289.1; -.
DR HSSP: P01556; ZCHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR ProDom: PD012805; Enterotoxin_B.1.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;
Query Match 100.0%; Score 62; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VEVPGSQHDSQ 12
DB 71 VEVPGSQHDSQ 82
RESULT 5
Q93V32 PRELIMINARY; PRT; 124 AA.
AC Q93V32;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.
GN LTH B SUBUNIT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1032 (ENTEROTOXIGENIC);
RA Komase K.;
RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-1032 (ENTEROTOXIGENIC);
RX MEDLINE=95091056; PubMed=7998417;
RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
RA Hattori N., Watanabe K., Suzuki Y., Negamine T., Aizawa C., Oya A.,
```



RA Kurata T.;  
 RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with  
 RT a trace amount of the holotoxin as an adjuvant for nasal influenza  
 RL vaccine.";  
 RL Vaccine 12:1083-1089(1994).  
 DR EMBL; AB011677; BAA25726.1; -  
 SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 62; DB 2; Length 124;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12  
 |||||  
 DB 71 VEPGSOHDSQ 82

RESULT 6  
 O94M01 PRELIMINARY; PRT; 124 AA.

AC O94M01;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DE CTXB.  
 GN CTXB.  
 OS Vibrio phage CTX.  
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxId=141904;  
 RN [1]  
 RP Bhattacharyya T., Nandy R.K., Nair G.B.;  
 RT "The entire core region of the ctx-phi (ctx-prophage) in VCE 232, an  
 RT environmental strain of V. cholerae";  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF414369; AAL09682.1; -  
 SQ SEQUENCE 124 AA; 13884 MW; 8D4C251591B16891 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 62; DB 9; Length 124;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VEPGSOHDSQ 12  
 |||||  
 DB 71 VEPGSOHDSQ 82

RESULT 7  
 O9NKD5 PRELIMINARY; PRT; 395 AA.

AC O9NKD5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DE HYPOTHETICAL 43.6 KDA PROTEIN.  
 GN BG:DS01514.3.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y, CN BW SP;  
 RA MEDLINE-9403001; PubMed-10471707;  
 RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,  
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
 RA Palazzolo M., Reese M.G., Spreading A., Tsang G., Wan K., Whitehead K.,  
 RA Celisner S., Rubin G.M.;  
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
 RT Drosophila melanogaster: the Adh region";  
 RL Genetics 153:179-219(1999).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y, CN BW SP;  
 RA Celisner S.E., Aghayani A., Arcaina T.T., Baxter E., Blazej R.G.,  
 RA Butenloft C., Champagne M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,  
 RA Houston K.A., Humasta S.R., Karia K., Kearney L., Kim E., Lee B.,  
 RA Lewis S., Li P., Lomostan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,  
 RA Nixon K., Pacled J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,  
 RA Sethi H., Sutr E., Svirskas R.R., Wan K.H., Weidburg T., Zhang R.,  
 RA Zieran L.L., Rubin G.M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003408; AAF44847.1; -  
 DR Flybase: FBgn0028907; BG:DS01514.3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 395 AA; 43561 MW; AE4F1C4ADD3DA73 CRC64;

Query Match  
 Best Local Similarity 72.7%; Score 43; DB 5; Length 395;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VEPGSOHDS 11  
 |||||  
 DB 330 VSWPGSTHIDA 340

RESULT 8  
 O947W6 PRELIMINARY; PRT; 802 AA.

AC O947W6;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DE PUTATIVE TNP-LIKE TRANSPOSABLE ELEMENT.  
 GN OSJNBA0029P06.9.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriaroidae; Oryzaeae; Oryza.  
 OX NCBI\_TaxId=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Saski C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC090873; AAK98653.1; -  
 SQ SEQUENCE 802 AA; 92934 MW; 01ADE54BA1CD1108 CRC64;

Query Match  
 Best Local Similarity 77.8%; Score 42; DB 10; Length 802;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 PGSOHDSE 12  
 |||||  
 DB 516 PGSOHDSE 524

RESULT 9  
 O94HM6 PRELIMINARY; PRT; 1109 AA.

AC O94HM6;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DE PUTATIVE TRANSPOSABLE ELEMENT.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriaroidae; Oryzaeae; Oryza.  
 OX NCBI\_TaxId=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-NIPPONBARE;  
 RA Spiegel L., de la Bastide M., Nascimento L., Kirchoff K., King L.,  
 RA Preston R., Vil M.D., Baker J., Bell M., Zutavern T., Santos L.,  
 RA Miller B., Kuit K., Rodriguez S., Cunniss D.M., Balija V., Shah R.,  
 RA Bahret A., Bal H., O'Shaughnessy A., Dedhia N., McCombie W.R.;  
 RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone  
 RT OSJNBa00890D5, from chromosome 10, complete sequence";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-NIPPONBARE;  
 RA McCombie W.R.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-NIPPONBARE;  
 RA McCombie W.R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-NIPPONBARE;  
 RA Palmer L.E., Spiegel L., de la Bastide M., Nascimento L., Kirchoff K.,  
 RA King L., Preston R., Vil M.D., Baker J., Bell M., Zutavern T.,  
 RA Santos L., Miller B., Kuit K., Rodriguez S., Cunniss D.M., Balija V.,  
 RA Shah R., Bahret A., Bal H., O'Shaughnessy A., Dedhia N.,  
 RA McCombie W.R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC078944; AAK92611.1; -  
 SQ SEQUENCE 1109 AA; 127116 MW; 5C026FEF70EDB85 CRC64;

Query Match 67.7%; Score 42; DB 10; Length 1109;  
 Best Local Similarity 77.8%; Pred. No. 21;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 PGSQHDSQ 12  
 |||||:|:  
 DB 526 PGSQHISE 534

RESULT 10  
 O94EB4  
 ID O94EB4 PRELIMINARY; PRT; 1352 AA.  
 AC O94EB4;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE P0002B05.25 (TREMblrel. 19, Last annotation update)  
 GN P0002B05.25.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriarthoidae; Oryzae; Oryza.  
 NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV, NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0002B05";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003141; BAB63505.1; -  
 SQ SEQUENCE 1352 AA; 154782 MW; DFF265A2B971B8F8 CRC64;

Query Match 67.7%; Score 42; DB 10; Length 1352;  
 Best Local Similarity 77.8%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 PGSQHDSQ 12  
 |||||:|:  
 DB 463 PGSQHISE 471  
 RESULT 11

O948C7  
 ID O948C7 PRELIMINARY; PRT; 2535 AA.  
 AC O948C7;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE P0002B05.25 (TREMblrel. 19, Last annotation update)  
 GN P0002B05.25.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriarthoidae; Oryzae; Oryza.  
 NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Sasaki C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC079852; AAL25171.1; -  
 SQ SEQUENCE 2535 AA; 289895 MW; DD280569D154168B CRC64;

Query Match 67.7%; Score 42; DB 10; Length 2535;  
 Best Local Similarity 77.8%; Pred. No. 51;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 PGSQHDSQ 12  
 |||||:|:  
 DB 526 PGSQHISE 534

RESULT 12  
 O86582  
 ID O86582 PRELIMINARY; PRT; 242 AA.  
 AC O86582;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DE P0002B05.25 (TREMblrel. 17, Last annotation update)  
 GN P0002B05.25.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetes.  
 NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Murphy L., Harris D.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Parkhill J., Barrett B.G., Rajandream M.A.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL MOL. Microbiol. 21:77-96(1996).  
 DR EMBL: AL031184; GAA20190.1; -  
 DR InterPro: IPR002502; Amidase\_2.  
 DR Pfam: PF01510; Amidase\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 242 AA; 26018 MW; 6DD10FF18A2EC544 CRC64;

Query Match 64.5%; Score 40; DB 2; Length 242;  
 Best Local Similarity 70.0%; Pred. No. 9.3;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VEVPQSCHID 10  
 |||||:|  
 DB 204 VEVPQSDHTD 213

## RESULT 13

031335 PRELIMINARY; PRT; 254 AA.  
 AC 031335:  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ATP-DEPENDENT NUCLEASE SUBUNIT A (FRAGMENT).  
 GN ADDA.  
 OS Bacillus cereus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1396;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 10987;  
 RX MEDLINE=99231848; PubMed=10217496;  
 RA Okstad O.A., Hegna I., Lindbaeck T., Rishovd A.L., Kolsto A.B.;  
 RT "Genome organisation is not conserved between Bacillus cereus and  
 RT Bacillus subtilis.";  
 RL Microbiology 145:621-631(1999).  
 DR EMBL: Y11217; CAN72103.1;  
 DR InterPro: IPR000212; UVRD-helicase.  
 DR Pfam: PF00580; UVRD-helicase; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 254 AA; 29212 MW; 24392E11338D99EA CRC64;

Query Match 64.5%; Score 40; DB 2; Length 254;  
 Best Local Similarity 58.3%; Pred. No. 9.8;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEVPQSCHID 12  
 |||||:|  
 DB 35 IDEPQSCHIRKQ 46

## RESULT 14

09NP06 PRELIMINARY; PRT; 192 AA.  
 AC 09NP06:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE LTP, LCAT-LIKE LYSOPHOSPHOLIPASE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Pluvinet R., Estivill X., Escarceller M., Sumoy L.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Aufray C., Ansoorge W., Ballabio A., Estivill X., Gibson K.,  
 RA Lehnach H., Poustka A., Lundberg J.;  
 RT "The European IMAGE consortium for Integrated Molecular analysis of  
 RT human gene transcripts.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL38957; CAB97531.1;  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 192 AA; 21609 MW; 04A7A8CB344F213 CRC64;

Query Match 62.9%; Score 39; DB 4; Length 192;  
 Best Local Similarity 66.7%; Pred. No. 11;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVPGSCHID 10  
 |||||:|  
 DB 166 ELPGSEHIE 174

## RESULT 15

09RJ28 PRELIMINARY; PRT; 201 AA.  
 AC 09RJ28:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE AMIDASE.  
 GN SCG11A.03.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL133210; CAB61584.1;  
 DR InterPro: IPR002502; Amidase\_2.  
 DR Pfam: PF01510; Amidase\_2; 1.  
 SQ SEQUENCE 201 AA; 22749 MW; B8EF477E06A20468 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 201;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVPGSCHID 10  
 |||||:|  
 DB 176 EVPGSDHTD 184

Search completed: October 24, 2002, 15:22:25  
 Job time : 17.2008 secs

**THIS PAGE BLANK (USPTO)**



xx Disclosure: Page 15; 62pp; English.

ps

xx The invention relates to peptide fragments of the Escherichia coli heat

xx labile enterotoxin (Etx) and its closely related homologue, cholera

cc toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous

cc GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

cc composed of one A subunit and five identical B subunits. The A subunit

cc is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

cc ribosyltransferase activity, while the B subunits (EtxB and CtxB)

cc facilitate the entry of subunit A into the host cell via the binding and

cc cross-linking of GM-1 receptors. Although GM-1 binding is responsible

cc for some of the effects of Etx and Ctx, it has been found that certain

cc effects of the toxins, such as immunomodulation, are not mediated

cc through GM-1 binding. The peptides of the invention are fragments of the

cc beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as

cc normal EtxB and CtxB subunits, except that they do not bind or cross link

cc GM-1. They may be used in medicine as an immunomodulator or adjuvant.

cc They may also be used as an inhibitor for toxin-induced diarrhoea.

cc Therefore, the peptides may be used in the production of a composition

cc for treating, preventing and/or modulating a disease associated with an

cc immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-187463

cc represent preferred peptides of the invention, AAY87460 being

cc particularly preferred.

cc

xx

xx Sequence 21 AA:

xx

xx

Query Match	100.0%	Score 106;	DB 21;	Length 21;
Best Local Similarity	100.0%;	Pred. No. 5.1e-11;		
Matches 21; Conservative	0;	Mismatches	0;	Gaps 0;

```
Qy 1 GATFQVEVPGSQHIDSQKAI 21
    |||||
Db 1 GATFQVEVPGSQHIDSQKAI 21
```

RESULT 2  
AAY87463

ID AAY87463 standard; peptide; 21 AA.

AC AAY87463;

DT 03-JUL-2000 (first entry)

DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.

KW Heat labile enterotoxin subunit B; EtXB;

KW adjuvant; immune disorder; diarrhoea.

OS *Escherichia coli*.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GB02970.

PR 07-SEP-1998; 98GB-0019484.

PA (UYBR-) UNIV BRISTOL.

PI Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

PT Derivatives of *Escherichia coli* heat labile enterotoxins useful as

PT glycolipid receptor GM-1 -

PS Disclosure; Page 15; 62pp; English.

CC The invention relates to peptide fragments of the Escherichia coli heat

CC labile enterotoxin (Etx) and its closely related homologue, cholera  
CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous  
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
CC composed of one A subunit and five identical B subunits. The A subunit  
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
CC facilitate the entry of subunit A into the host cell via the binding and  
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
CC for some of the effects of Etx and Ctx, it has been found that certain  
CC effects of the toxins, such as immunomodulation, are not mediated  
CC through GM-1 binding. The peptides of the invention are fragments of the  
CC beta-1- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
CC Therefore, the peptides may be used in the production of a composition  
CC for treating, preventing and/or modulating a disease associated with an  
CC immune disorder and/or toxin-induced diarrhoea. Sequences AY87460-AY87463  
CC represent preferred peptides of the invention, AY87460 being  
CC particularly preferred.  
XX  
XX  
Sequence 21 AA;

Query Match	95.3%	Score 101	DB 21	Length 21
Best Local Similarity	95.28%	Pred. No. 3.6e-10		
Matches 20, Conservative		0; Mismatches 1;	Indels 0;	Gaps 0

```
QY 1 GATFOVEVPGSQHIDSQKAI 21
    1 | | | | | | | | | | | | | |
Db 1 GETFOVEVPGSQHIDSQKAI 21
```

RESULT 3  
AAR72545

AAAR  
AC  
XX

DT 28-NOV-1995 (first entry)

DE ADP-ribose

KW ADP-ribosylating toxin; pertussis holotoxin; B-

KW Bordetella pertussis vaccines.

05 Bacteria sp.

PN EP646599-A.

PD 05-APR-1995

PF 23-AUG-1994

PR 24-AUG-1993; 93US-0110947.

XX

PA (UYAL-) UNIV ALBERTA.

PI     Armstrong GD,     Cockle SA,

XX

XX

PT crystalline forms of pertussis holotoxin and its complexes

XX

XX

CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat







XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 KW Vibrio cholerae.  
 OS EP95426-A.  
 XX 30-NOV-1983.  
 PD 26-MAY-1983; 83EP-0401052.  
 XX 26-MAY-1983; 82FR-0009167.  
 PR (CNRS ) CNRS CENT NAF RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 XX WPI: 1983-834645/49.  
 DR Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX Claim 8; Page 11; 13pp; French.  
 PS The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX Sequence 47 AA:  
 SQ Query Match 87.3%; Score 92.5; DB 4; Length 47;  
 Best Local Similarity 90.9%; Pred. No. 2.5e-08;  
 Matches 20; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
 OY 1 GATFQ-VEVPGSQHIDSOKKAI 21  
 DB 16 GATFEVVEVPGSQHIDSOKKAI 37  
 RESULT 9  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 AC AAP50436;  
 XX AAP50436;  
 DT 01-JAN-1980 (first entry)  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX Synthetic.  
 OS W08502611-A.  
 XX W08502611-A.  
 PD 20-JUN-1985.  
 XX 12-DEC-1984; 84WO-US02030.  
 PR 12-DEC-1983; 83US-0559469.  
 XX 12-DEC-1983; 83US-0559469.  
 PA (SCRI-) SCRIPPS CLINIC RES.

XX Houghten RA;  
 PI WPI: 1985-159230/26.  
 DR New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX Claim 8; Page 100; 120pp; English.  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX Sequence 46 AA:  
 SQ Query Match 84.0%; Score 89; DB 6; Length 46;  
 Best Local Similarity 94.4%; Pred. No. 9.6e-08;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GATFQVEVPGSQHIDSOK 18  
 DB 11 GATFQVEVPGSQHIDSOK 28  
 RESULT 10  
 AAP50439  
 ID AAP50439 standard; protein; 41 AA.  
 AC AAP50439;  
 XX AAP50439;  
 DT 01-JAN-1980 (first entry)  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX Synthetic.  
 OS W08502611-A.  
 XX W08502611-A.  
 PD 20-JUN-1985.  
 XX 12-DEC-1984; 84WO-US02030.  
 PR 12-DEC-1983; 83US-0559469.  
 XX 12-DEC-1983; 83US-0559469.  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX Houghten RA;  
 DR WPI: 1985-159230/26.  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX Claim 8; Page 100; 120pp; English.  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX Sequence 41 AA:  
 SQ

Query Match	83.0%	Score 88;	DB 6;	Length 41;
Best Local Similarity	94.4%;	Pred. No. 1.2e-07;		
Matches	17;	Conservative	0;	Mismatches 1; Indels 0;
QY	1	GATFOVEVPGSOHDSOK	18	
Db	9	GATFOVEVPGSOHDSOK	26	

RESULT 11	
AAP30265	
ID	AAP30265 standard; Protein; 26 AA.
XX	
AC	AAP30265;
XX	
DT	21-APR-1992 (first entry)
XX	
DE	Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
DE	carries an Arg at posns. 67 and 73.
XX	
KW	Cholera vaccine; therapy; E.coli Infection; enterotoxin LT.
XX	
OS	Vibrio cholerae.
XX	

RESULT 12	
AAFP93498	
ID	AAFP93498 standard; protein; 15 AA
XX	
AC	AAFP93498;

XX	03-MAY-1990 (first entry)	
DT		
XX		
DE	CTP3 epitope of the Cholera toxin B subunit.	
XX		
KM	CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein.	
KN	vaccine; immunotherapy; ds;	

RESULT 13	
AAR85125	
ID	AAR85125 standard; peptide; 15 AA.
XX	
AC	AAR85125;
XX	
DT	13-JUN-1996 (first entry)
XX	
DE	Cholera toxin B antigenic peptide fragment CTP3.
XX	
KW	Conjugate; cholera; B toxin; peptide fragment; microparticulate;
KW	inert carrier; modified silica; thyroglobulin; oral vaccine;
KW	immunisation; infection; insoluble; digestive tract; antigen;
KW	intestines; antipodes; secretory; IGA class.
XX	
OS	Vibrio cholerae.
XX	
PN	W09529701-A1.
XX	
PD	09-NOV-1995.
XX	
PF	02-MAY-1995; 95WO-EP01661.
XX	
PR	03-MAY-1994; 94IL-0109519.
XX	
PA	(YEDA ) YEDA RES & DEV CO LTD.

XX Marks RS, Mirelman D, Sela M;  
 XX  
 XX WPI; 1995-403805/51.  
 DR  
 XX  
 PT Vaccines for oral immunisation against infecting agents, e.g.  
 PT cholera - comprise a conjugate of an antigen of an infecting agent  
 PT covalently bound to micro:particulate inert carrier, e.g. modified  
 PT aldehyde silica  
 XX  
 PS Claim 7; Page 25; 40pp; English.  
 XX  
 CC A compsn. comprising a conjugate of an antigenic cholera B toxin  
 CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently  
 CC bound to a microparticulate inert carrier (e.g. modified silica or  
 CC thyroglobulin) can be used as an oral vaccine for immunisation  
 CC against cholera infection. The inert carrier is insoluble in the  
 CC digestive tract, allowing presentation of the antigen in the  
 CC intestines, where it will elicit antibodies mainly of the  
 CC secretory IgA class.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 Query Match 71.7%; Score 76; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 VEPGSOHDSOKKA 20  
 ||||||||||||  
 Db 1 VEPGSOHDSOKKA 15

RESULT 14  
 AAR76748  
 ID AAR76748 standard; Protein: 23 AA.  
 XX  
 AC AAR76748;  
 XX  
 DT 18-MAR-1996 (first entry)  
 XX  
 DE Residues 50-64 of cholera toxin B subunit and FimH 224-226.  
 XX  
 KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;  
 KW FimH; FimH; FimH; receptor binding site; PCR; amplify; ss.  
 XX  
 OS Chimeric - Vibrio cholerae.  
 OS Chimeric - Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1  
 FT Peptide 2..4 /note= "Represents FimH residue 224"  
 FT Peptide 5..19 /note= "Linker peptide"  
 FT Peptide /note= "Cholera toxin B subunit 50-64"  
 FT Peptide 20..22 /note= "Linker peptide"  
 FT Misc-difference 23 /note= "Represents FimH residue 226"  
 FT  
 XX  
 PN WO9520657-A1.  
 XX  
 PD 03-AUG-1995.  
 XX  
 PP 27-JAN-1995; 95WO-DK00042.  
 XX  
 PR 27-JAN-1994; 94US-0187166.  
 XX  
 XX (GBR-) GX BIOSYSTEMS AS.  
 PA  
 XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;  
 PI  
 XX WPI; 1995-275442/36.  
 DR

DR N-PSDB; AAQ93061.  
 XX  
 PT Receptor specific bacterial adhesins - useful for targetting active  
 PT compounds and microbial cells to locations of receptors  
 PT  
 XX  
 PS Example 3; Page 58; 152pp; English.  
 XX  
 CC This sequence is encoded by a fragment of the the plasmid p19A93  
 CC which was used in the production of fimH fusion genes comprising  
 CC the cholera toxin B subunit inserted into the fimH gene. This insert  
 CC shows the inclusion of the B subunit into the fimH protein at position  
 CC 224-226. The chimeric genes were then opt. further modified by insertion  
 CC of the hepatitis B virus surface antigen pre-S2 region into a different  
 CC position of the fimH adhesin of type 1 fimbriae. Restriction site handles  
 CC (BglII-sites) were introduced into the fimH gene, and the foreign  
 CC epitopes are then inserted in-frame. In the selected positions the  
 CC insertion of the epitopes did not significantly alter the adhesive  
 CC function of the fimH protein. The expression of the chimeric proteins  
 CC on the surface of fimbriae on bacterial hosts illustrated the possibility  
 CC of using bacterial adhesins as general presenters of foreign antigens and  
 CC epitopes. These chimeric genes may be used in the production of variant  
 CC FimH adhesins which may be useful for targetting active compounds  
 CC and microbial cells to locations comprising selected receptors to which  
 CC the adhesins bind.  
 CC  
 XX  
 SQ Sequence 23 AA;  
 Query Match 71.7%; Score 76; DB 16; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 VEPGSOHDSOKKA 20  
 ||||||||||||  
 Db 5 VEPGSOHDSOKKA 19

RESULT 15  
 AAY87461  
 ID AAY87461 standard; peptide: 12 AA.  
 XX  
 AC AAY87461;  
 XX  
 DT 03-JUL-2000 (first entry)  
 XX  
 DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.  
 XX  
 KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
 KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
 KW adjuvant; immune disorder; diarrhoea.  
 XX  
 OS Vibrio cholerae.  
 OS Escherichia coli.  
 XX  
 PN WO200014114-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PP 07-SEP-1999; 99WO-GB02970.  
 XX  
 PR 07-SEP-1998; 98GB-0019484.  
 XX  
 XX (UYBR-) UNIV BRISTOL.  
 PA  
 XX Williams NA, Hirst TR;  
 PI  
 XX WPI; 2000-256943/22.  
 DR  
 XX  
 XX Derivatives of Escherichia coli heat labile enterotoxins useful as  
 PT immunomodulators and for treating diarrhea and which do not bind the  
 PT glycolipid receptor GM-1 -  
 XX  
 PS Disclosure; Page 15; 62pp; English.  
 XX

CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4' alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AY87460-187463  
 CC represent preferred peptides of the invention, AY87460 being  
 CC particularly preferred.

XX  
 SQ Sequence 12 AA;

Query Match 58.5%; Score 62; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.00075;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VEYVGSQHIDSQ 17  
 DB 1 VEYVGSQHIDSQ 12

Search completed: October 24, 2002, 15:35:04  
 Job time : 34.7049 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:33:23 : Search time 12.0492 Seconds  
(without alignments)  
42.570 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATGVEVPSQHDSQKRAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 177902

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCtus.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	95.3	93	2	US-08-292-968-26 Sequence 26, Appl
2	101	95.3	93	2	US-08-467-974-26 Sequence 26, Appl
3	101	95.3	93	2	US-08-467-536-26 Sequence 26, Appl
4	101	95.3	93	3	US-08-467-976-26 Sequence 26, Appl
5	101	95.3	93	4	US-09-082-514-26 Sequence 26, Appl
6	37	34.9	75	1	US-08-350-884-35 Sequence 35, Appl
7	37	34.9	75	1	US-08-709-173-35 Sequence 35, Appl
8	37	34.9	75	2	US-08-709-173-35 Sequence 35, Appl
9	37	34.9	75	2	US-08-833-678A-3 Sequence 3, Appl
10	37	34.9	75	4	US-08-529-169A-3 Sequence 3, Appl
11	37	34.9	76	1	US-08-232-815-2 Sequence 2, Appl
12	37	34.9	76	1	US-08-350-906-2 Sequence 2, Appl
13	37	34.9	76	1	US-08-468-274-2 Sequence 2, Appl
14	37	34.9	76	5	PCT-US95-04536-2 Sequence 2, Appl
15	36	34.0	27	3	US-09-045-632-83 Sequence 83, Appl
16	36	34.0	38	2	US-08-726-306A-101 Sequence 101, App
17	36	34.0	77	6	US-08-450-834-6 Sequence 6, Appl
18	36	34.0	76	1	Patent No. 5510474 Sequence 94, Appl
19	36	34.0	78	2	US-08-505-486-94 Sequence 94, Appl
20	36	34.0	78	3	US-08-801-028-94 Sequence 94, Appl
21	36	34.0	78	3	US-09-340-154-94 Sequence 94, Appl
22	36	34.0	78	5	PCT-US95-09338-94 Sequence 94, Appl
23	36	34.0	78	5	PCT-US95-09338-94 Sequence 94, Appl
24	36	34.0	79	3	US-09-100-804-23 Sequence 23, Appl
25	36	34.0	93	4	US-09-069-023-15 Sequence 15, Appl
26	33	31.1	21	1	US-08-355-888A-14 Sequence 14, Appl
27	33	31.1	21	2	US-08-693-697-14 Sequence 14, Appl

28	33	31.1	21	3	US-08-693-696-14	Sequence 14, Appl
29	32	30.2	41	6	5208144-18	Patent No. 5208144
30	31.5	29.7	35	1	US-08-418-893D-16	Sequence 16, Appl
31	31	29.2	28	1	US-08-255-558B-6	Sequence 6, Appl
32	31	29.2	28	3	US-09-045-632-61	Sequence 61, Appl
33	31	29.2	60	1	US-08-197-792-11	Sequence 11, Appl
34	31	29.2	60	1	US-08-459-850-11	Sequence 11, Appl
35	31	29.2	60	1	US-08-459-214-11	Sequence 11, Appl
36	31	29.2	75	2	US-08-332-562A-29	Sequence 29, Appl
37	31	29.2	76	2	US-08-284-351B-53	Sequence 53, Appl
38	31	29.2	76	4	US-09-218-950-53	Sequence 53, Appl
39	31	29.2	76	4	US-09-180-827-3	Sequence 3, Appl
40	31	29.2	93	1	US-08-158-189-16	Sequence 16, Appl
41	31	29.2	95	1	US-07-956-700B-23	Sequence 23, Appl
42	31	29.2	95	1	US-08-476-537-23	Sequence 23, Appl
43	31	29.2	95	1	US-08-485-607-23	Sequence 23, Appl
44	31	29.2	95	2	US-08-475-879-23	Sequence 23, Appl
45	31	29.2	98	3	US-09-045-632-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 3856122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COHEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-292-968-26

Query Match 95.3%; Score 101; DB 2; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21  
Db 35 GETFOVEVPGSHIDSOKKAI 55

RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385

; GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,974

FILING DATE:

CLASSIFICATION: 435

APPLICATION NUMBER: US 08/467,536

FILING DATE: 06-JUN-1995

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-454 MIS.vg

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-974-26

Query Match 95.3%; Score 101; DB 2; Length 93;

Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21

Db 35 GETFOVEVPGSHIDSOKKAI 55

RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304

; GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.

APPLICANT: HAZES, Bart

TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,536

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

APPLICATION NUMBER: US 08/292,968

FILING DATE: 22-AUG-1994

APPLICATION NUMBER: US 08/251,121

FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/110,947

FILING DATE: 24-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: STEWART, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-455 MIS.vg

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-467-536-26

Query Match 95.3%; Score 101; DB 2; Length 93;

Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSHIDSOKKAI 21

Db 35 GETFOVEVPGSHIDSOKKAI 55

RESULT 4

US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022

; GENERAL INFORMATION:

```

; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,976
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 22-AUG-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/251,121
; FILING DATE: 31-MAY-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-976-26

Query Match          95.3%; Score 101; DB 3; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 5
US-09-082-514-26
; Sequence 26, Application US/09082514
; Patent No. 6168928
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,514
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,968
; FILING DATE: 24-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-082-514-26

Query Match          95.3%; Score 101; DB 4; Length 93;
Best Local Similarity 95.2%; Pred. No. 1.7e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 6
US-08-350-884-35
; Sequence 35, Application US/08350884
; Patent No. 5585258
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI LIM
; APPLICANT: KOO, GEORGE
; TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/350,884
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/680,296
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:

```

NAME: CIOTTI, THOMAS E.  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20100.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-884-35

Query Match 34.9%; Score 37; DB 1; Length 75;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
QY 1 GATFOVEVPGSQHDSOKKAI 21  
| | : | | : | | : | |  
Db 9 GKTTTLEVSSDTIDNVKSKI 29

RESULT 7  
US-08-709-173-35  
Sequence 35, Application US/08709173  
Patent No. 5712145  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, OUI LIM  
APPLICANT: KUO, GEORGE  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,173  
FILING DATE: 06-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/680,296  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: CIOTTI, THOMAS E.  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20100.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-709-173-35

Query Match 34.9%; Score 37; DB 1; Length 75;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSOKKAI 21  
| | : | | : | | : | |  
Db 9 GKTTTLEVSSDTIDNVKSKI 29

RESULT 8  
US-08-709-177-35  
Sequence 35, Application US/08709177  
Patent No. 5885799  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, OUI LIM  
APPLICANT: KUO, GEORGE  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,177  
FILING DATE: 06-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/680,296  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: CIOTTI, THOMAS E.  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20100.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-709-177-35

Query Match 34.9%; Score 37; DB 2; Length 75;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSOKKAI 21  
| | : | | : | | : | |  
Db 9 GKTTTLEVSSDTIDNVKSKI 29

RESULT 9  
US-08-833-678A-3  
Sequence 3, Application US/08833678A  
Patent No. 5989905  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, OUI LIM  
APPLICANT: HAN, JANG  
APPLICANT: CHOE, JOONHO  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION



STREET: Intellectual Property - R440, P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/833,678A  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/529,169  
FILING DATE: 15-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Harbin, Alisa A.  
REGISTRATION NUMBER: 33,895  
REFERENCE/DOCKET NUMBER: 0100.005  
TELEPHONE: (510) 923-3274  
TELEFAX: (510) 655-3542  
TELEX: n/a  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-833-678A-3

Query Match 34.9%; Score 37; DB 2; Length 75;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
DB 9 GKTTLEVESSDTIDNVKSKI 29

RESULT 10  
US-08-529-169A-3  
Sequence 3, Application US/08529169A  
Patent No. 6194140  
GENERAL INFORMATION:  
APPLICANT: Houghton, Michael  
APPLICANT: Choo, Qui-Lim  
APPLICANT: Han, Jang  
APPLICANT: Choe, Joonho  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: Intellectual Property - R440, P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,169A  
FILING DATE: 15-SEP-1995  
CLASSIFICATION: 4325  
ATTORNEY/AGENT INFORMATION:  
NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895  
REFERENCE/DOCKET NUMBER: 0100.005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-3274  
TELEFAX: (510) 655-3542  
TELEX: n/a  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-529-169A-3

Query Match 34.9%; Score 37; DB 4; Length 75;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
DB 9 GKTTLEVESSDTIDNVKSKI 29

RESULT 11  
US-08-232-815-2  
Sequence 2, Application US/08232815  
Patent No. 5503977  
GENERAL INFORMATION:  
APPLICANT: Johansson, Nils  
APPLICANT: Varshavsky, Alexander  
TITLE OF INVENTION: SPLIT UBIQUITIN PROTEIN SENSOR  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kevin M. Farrell  
STREET: P.O. Box 999  
CITY: York Harbor  
STATE: Maine  
COUNTRY: USA  
ZIP: 03911  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,815  
FILING DATE: 22-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: CIT-9303  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 207 363-0558  
TELEFAX: 207 363-0528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-232-815-2

Query Match 34.9%; Score 37; DB 1; Length 76;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
DB 10 GKTTLEVESSDTIDNVKSKI 30

RESULT 12  
US-08-350-906-2  
; Sequence 2, Application US/08350906  
; Patent No. 5585245  
; GENERAL INFORMATION:  
; APPLICANT: Johnsson, Nils  
; APPLICANT: Varshavsky, Alexander  
; TITLE OF INVENTION: UBIQUITIN-BASED SPLIT PROTEIN SENSOR  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: Maine  
; COUNTRY: USA  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/350,906  
; FILING DATE: 07-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-9303A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 207 363-0558  
; TELEFAX: 207 363-0528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 76 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-350-906-2

Query Match 34.9%; Score 37; DB 1; Length 76;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKAI 21  
| | : | | | | : | |  
Db 10 GKTTLEVSDDTDNVKSKI 30

RESULT 13  
US-08-468-274-2  
; Sequence 2, Application US/08468274  
; Patent No. 5714464  
; GENERAL INFORMATION:  
; APPLICANT: Piraino, Frank and Brandt, Curtis R  
; TITLE OF INVENTION: Anti-Viral Mushroom Extracts  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nicholas J. Seay  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,274  
; FILING DATE:  
; CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608/251-5000  
; TELEFAX: 608/251-9166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 76 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
; ORGANISM: Rozites caperata  
US-08-468-274-2

Query Match 34.9%; Score 37; DB 1; Length 76;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSQKAI 21  
| | : | | | | : | |  
Db 10 GKTTLEVSDDTDNVKSKI 30

RESULT 14  
PCT-US95-04536-2  
; Sequence 2, Application PC/TUS9504536  
; GENERAL INFORMATION:  
; APPLICANT: California Institute of Technology  
; TITLE OF INVENTION: UBIQUITIN-BASED SPLIT PROTEIN SENSOR  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kevin M. Farrell  
; STREET: P.O. Box 999  
; CITY: York Harbor  
; STATE: Maine  
; COUNTRY: USA  
; ZIP: 03911  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04536  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/350,906  
; FILING DATE: 07-DEC-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farrell, Kevin M.  
; REGISTRATION NUMBER: 35,505  
; REFERENCE/DOCKET NUMBER: CIT-9303A WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 207 363-0558  
; TELEFAX: 207 363-0528  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 76 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-04536-2

Query Match 34.9%; Score 37; DB 5; Length 76;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFQVEVPGSOHDSOKKAI 21  
 1 1 1 1 1 1 1 1 1 1  
 DB 10 GKTITLEVSSDPTIDNVKSKI 30

RESULT 15

US-09-045-632-83  
 ; Sequence 83, Application US/09045632  
 ; Patent No. 6001575  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HUGANIR, Richard L.  
 ; APPLICANT: Dong, Hualing  
 ; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND  
 ; TITLE OF INVENTION: GRIP-RELATED MOLECULES  
 ; NUMBER OF SEQUENCES: 105  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
 ; STREET: 130 Water Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/045,632  
 ; FILING DATE: 19-MAR-1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/041,016  
 ; FILING DATE: 19-MAR-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Corless, Peter F.  
 ; REGISTRATION NUMBER: 33,860  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-523-3400  
 ; TELEFAX: 617-523-6440  
 ; INFORMATION FOR SEQ ID NO: 83:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 27 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-09-045-632-83

Query Match 34.0%; Score 36; DB 3; Length 27;  
 Best Local Similarity 46.7%; Pred. No. 8.8;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 6 VEPGSGHDSOKKA 20  
 1 1 1 1 1 1 1 1 1 1  
 DB 2 VDVPHASAVDALKKA 16

Search completed: October 24, 2002, 15:38:24  
 Job time : 14.0492 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:09 ; Search time 139.77 Seconds  
(without alignments)  
52.884 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106  
Sequence: 1 GATFOVEVPGSQHDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 2671256

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCrUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	21	US-09-786-648-4	Sequence 4, Appli
2	101	95.3	21	US-09-786-648-5	Sequence 10, Appli
3	101	95.3	93	US-08-110-947-10	Sequence 26, Appli
4	101	95.3	93	US-08-110-947A-26	Sequence 26, Appli
5	101	95.3	93	US-08-251-121-26	Sequence 1, Appli
6	76	71.7	15	US-08-732-371-1	Sequence 1, Appli
7	76	71.7	15	US-08-732-371A-1	Sequence 1, Appli

8	62	58.5	12	21	US-09-786-648-3	Sequence 3, Appli
9	46	43.4	99	21	US-09-758-463-761	Sequence 761, App
10	44	41.5	84	26	US-60-188-162-4087	Sequence 4087, App
11	44	41.5	91	21	US-09-708-427-6862	Sequence 6862, App
12	42	39.6	51	22	US-09-866-066-35	Sequence 35, Appli
13	41	38.7	45	1	PCT-US01-01334-5166	Sequence 5166, App
14	41	38.7	45	21	US-09-764-874-5166	Sequence 5166, App
15	40.5	38.2	44	26	US-60-143-993-1325	Sequence 1325, App
16	40	37.7	91	19	US-09-595-298A-470	Sequence 470, App
17	39	36.8	7	21	US-09-786-648-2	Sequence 2, Appli
18	39	36.8	18	14	US-09-072-955A-27	Sequence 27, Appli
19	39	36.8	61	21	US-09-708-427-44143	Sequence 44143, A
20	39	36.8	76	21	US-09-708-427-44142	Sequence 44142, A
21	39	36.8	97	23	US-09-902-540-12927	Sequence 12927, A
22	38	35.8	51	1	PCT-US01-00663-37905	Sequence 37905, A
23	38	35.8	51	22	US-09-864-761-41979	Sequence 41979, A
24	38	35.8	58	26	US-60-196-710-4922	Sequence 4922, App
25	38	35.8	81	18	US-09-417-507-30924	Sequence 30924, App
26	38	35.8	83	20	US-09-673-395A-328	Sequence 328, App
27	38	35.8	90	22	US-09-834-366-26047	Sequence 26047, A
28	38	35.8	90	26	US-60-197-873-26047	Sequence 26047, A
29	38	35.8	94	20	US-09-621-976-7007	Sequence 7007, App
30	38	35.8	94	26	US-60-147-459-7007	Sequence 7007, App
31	37.5	35.4	94	22	US-09-834-366-17872	Sequence 17872, A
32	37.5	35.4	94	26	US-60-197-873-17872	Sequence 17872, A
33	37	34.9	15	1	PCT-US00-24802-3025	Sequence 3025, App
34	37	34.9	15	1	PCT-US00-24802-3183	Sequence 3183, App
35	37	34.9	15	16	US-09-239-043-2176	Sequence 2176, App
36	37	34.9	15	16	US-09-239-043-2225	Sequence 2225, App
37	37	34.9	36	26	US-60-170-373-2676	Sequence 2676, App
38	37	34.9	36	26	US-60-181-996-1157	Sequence 1157, App
39	37	34.9	40	21	US-09-758-440-783	Sequence 783, App
40	37	34.9	61	19	US-09-593-110-3458	Sequence 3458, App
41	37	34.9	62	1	PCT-US00-05882-1475	Sequence 1475, App
42	37	34.9	62	23	US-09-925-301-1475	Sequence 1475, App
43	37	34.9	64	1	PCT-US01-01354-18194	Sequence 18194, A
44	37	34.9	64	21	US-09-764-905-18194	Sequence 18194, A
45	37	34.9	64	24	US-10-092-399-18194	Sequence 18194, A

#### ALIGNMENTS

RESULT 1

US-09-786-648-4

Sequence 4, Application US/09786648

GENERAL INFORMATION:

APPLICANT: Williams, Neil Andrew

TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci

TITLE OF INVENTION: Adjuvants

FILE REFERENCE: 7438

CURRENT APPLICATION NUMBER: US/09/786, 648

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: PCT/GB99/02970

PRIOR FILING DATE: 1999-09-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: MS DOS

SEQ ID NO 4

LENGTH: 21

TYPE: PRT

ORGANISM: E. COLI

FEATURE:

LOCATION: 45...65

OTHER INFORMATION: Isolated or synthetic EtxB beta4-alpha2 loop fragment derivabl

OTHER INFORMATION: human variant E. coli

US-09-786-648-4

Query Match 100.0%, Score 106, DB 21, Length 21;  
Best Local Similarity 100.0%, Pred. No. 1.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHDSOKKAI 21

```
Db      1 GATFOVEVPGSHIDSOKKAI 21
|||||
RESULT 2
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45..65
; OTHER INFORMATION: isolated or synthetic EtXB beta4-alpha2 loop fragment derivable f
US-09-786-648-5

Query Match      95.3%; Score 101; DB 21; Length 21;
Best Local Similarity 95.2%; Pred. No. 9.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATFOVEVPGSHIDSOKKAI 21
| |||||
Db      1 GATFOVEVPGSHIDSOKKAI 21

RESULT 3
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Hwy.
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
```

```
; TELLEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947-10

Query Match      95.3%; Score 101; DB 5; Length 93;
Best Local Similarity 95.2%; Pred. No. 6.2e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATFOVEVPGSHIDSOKKAI 21
| |||||
Db      35 GATFOVEVPGSHIDSOKKAI 55

RESULT 4
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COOMEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELLEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947A-26

Query Match      95.3%; Score 101; DB 5; Length 93;
Best Local Similarity 95.2%; Pred. No. 6.2e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATFOVEVPGSHIDSOKKAI 21
| |||||
Db      35 GATFOVEVPGSHIDSOKKAI 55

RESULT 5
US-08-251-121-26
```

```
; Sequence 26, Application US/08251121
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J.
; APPLICANT: STEIN, Penelope E.
; APPLICANT: COCKLE, Stephen A.
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; APPLICANT: ARMSTRONG, Glen D.
; APPLICANT: HAZES, Bart
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,121
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/110,947
; FILING DATE: 24-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-251-121-26
;
; Query Match 95.3%; Score 101; DB 6; Length 93;
; Best Local Similarity 95.2%; Pred. No. 6.2e-09;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 GATFOVEVPGSQHIDSOKKAI 21
; DB 35 GETFOVEVPGSQHIDSOKKAI 55
;
; RESULT 6
; US-08-732-371-1
; Sequence 1, Application US/08732371
; GENERAL INFORMATION:
; APPLICANT: MIRELMAN, David
; APPLICANT: MARKS, Robert S.
; APPLICANT: SELA, Michael
; TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,371
; FILING DATE: 09-JAN-1997
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 109519
; FILING DATE: 03-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: MIRELMAN-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-732-371-1
;
; Query Match 71.7%; Score 76; DB 11; Length 15;
; Best Local Similarity 100.0%; Pred. No. 1.2e-05;
; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 6 VEVPGSQHIDSOKKA 20
; DB 1 VEVPGSQHIDSOKKA 15
;
; RESULT 7
; US-08-732-371A-1
; Sequence 1, Application US/08732371A
; GENERAL INFORMATION:
; APPLICANT: MIRELMAN, David
; APPLICANT: MARKS, Robert S.
; APPLICANT: SELA, Michael
; TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,371A
; FILING DATE: 09-JAN-1997
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: IL 109519
; FILING DATE: 03-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: MIRELMAN-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
```

```

;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-732-371A-1

```

Query Match	71.7%;	Score 76;	DB 11;	length 15;
Best Local Similarity	100.0%;	Pred. No.	1.2e-05;	
Matches	15;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Qy	6	VEVPGSQHIDSQKKA	20
Db	1	VEVPGSQHIDSQKKA	15

```

1 RESULT 8
2 US-09-786-648-3
3
4 : Sequence 3, Application US/097866648
5
6 : GENERAL INFORMATION:
7
8 : APPLICANT: Williams, Neil Andrew
9
10 : TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
11
12 : TITLE OF INVENTION: Adjuvants
13
14 : FILE REFERENCE: 7438
15
16 : CURRENT APPLICATION NUMBER: US/09/786,648
17
18 : CURRENT FILING DATE: 2001-03-07
19
20 : PRIOR APPLICATION NUMBER: PCT/GB99/02970
21
22 : PRIOR FILING DATE: 1999-09-07
23
24 : NUMBER OF SEQ ID NOS: 6
25
26 : SOFTWARE: MS DOS
27
28 : SEQ ID NO 3
29
30 : LENGTH: 12
31
32 : TYPE: PRT
33
34 : ORGANISM: E. coli
35
36 : FEATURE:
37
38 : LOCATION: 50...61
39
40 : OTHER INFORMATION: Isolated or synthetic EtxB beta4-alpha2 loop fragment derivable
41
42 : OTHER INFORMATION: human variant E. coli
43
44 : US-09-786-648-3

```

Query Match	58.5%	Score 62;	DB 21;	Length 12;
Best Local Similarity	100.0%	Pred. No.	0.0022;	
Matches 12;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	6	VEVPGSQHIDSQ	17
Db	1	VEVPGSQHIDSQ	12

```

RESULT 9
US-09-758-463-761
: Sequence 761, Application US/09758463
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PM011
: CURRENT APPLICATION NUMBER: US/09/758,463
: CURRENT FILING DATE: 2001-01-11
: PRIOR APPLICATION NUMBER: 60/179,065
: PRIOR FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: 60/180,628
: PRIOR FILING DATE: 2000-02-04
: NUMBER OF SEQ ID NOS: 1304
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 761
: LENGTH: 99
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (9)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-463-761

```

Query Match	43.4%	Score 46:	DB 21;	Length 99;
Best Local Similarity	52.4%	Pred. No. 18:		
Matches 11: Conservative	2;	Mismatches	8;	Indels 0;
		Gaps	0;	
Qy	1	GATFOVEVPSGSHIDSKKAI	21	
Db	74	GTFVVGAGSSBEGDESGKAI	94	

```

OY      1 GATFOVEVPGSQHIDSQKAI 21
         | | | | | : | | | |
Db      74 GTTFQVGAGSSEGEDSEKAI 94

```

```

RESULT 10
US-60-188-162-4087
: Sequence 4087, Application US/60188162
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEIN,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: CL000518
: CURRENT APPLICATION NUMBER: US/60/188,162
: CURRENT FILING DATE: 2000-05-09
: NUMBER OF SEQ ID NOS: 5094
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4087
: LENGTH: 84
: TYPE: PRT
: ORGANISM: HUMAN
: US-60-188-162-4087

```

Query Match	41.5%	Score 44	DB 26	Length 84
Best Local Similarity	53.3%	Pred. No. 32		
Matches	8	Conservative	2	Mismatches 5
				Indels 0
				Gaps 0

```

QY      1 GATFQVEVPGSQHID 15
         |:|:||||| |
Db      42 GSTSRTEVPGQSHHD 56

```

```

RESULT 11
US-09-708-427-6862
: Sequence 6862, Application US/09708427
: GENERAL INFORMATION:
: APPLICANT: N. ALEXANDROV et al.
: TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
: TITLE OF INVENTION: THEREBY
: FILE REFERENCE: 2750-1243P
: CURRENT APPLICATION NUMBER: US/09/708,427
: CURRENT FILING DATE: 2000-11-09
: NUMBER OF SEQ ID NOS: 8364
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6862
: LENGTH: 91
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..91
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: misc_feature
: LOCATION: 1..91
: OTHER INFORMATION: Ceres Seq. ID 1816431
US-09-708-427-6862

```

Query Match	41.5%	Score 44	DB 21	Length 91
Best Local Similarity	47.6%	Pred. NO	35	
Matches 10; Conservative	2	Mismatches	9	Indels 0; Gaps 0

```
QY      1 GATFQVEVPGSQHIDSQKAI 21
          | | | : | | | : |
Db      58 GKTENLEVKGSEIIQYKNNI 78
```

RESULT 12  
US-09-866-066-35



```

; Sequence 35, Application US/09866066
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Christopher
; APPLICANT: Roberds, Steve
; APPLICANT: Ruble, Cara
; APPLICANT: Golow, Lisa
; APPLICANT: Karnovsky, Alla
; TITLE OF INVENTION: Human Ion Channels
; FILE REFERENCE: 00229, US1
; CURRENT APPLICATION NUMBER: US/09/866,066
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/207,152
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,257
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,119
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 35
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
us-09-866-066-35

Query Match          39.6%; Score 42; DB 22; Length 51;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy      2  ATGQVEPGSOHID 15
Db      3  SAFPTETPGSSHMD 16

RESULT 13
PCT-US01-01334-5166
; Sequence 5166, Application PC/TUS0101334
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC010PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01334
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14

```

```

; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787

```

;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,474  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/246,532  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/249,216  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,210  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/226,681  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,759  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/225,213  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/227,182  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,214  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/235,836  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/230,438  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/215,135  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: 60/225,266  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,081  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,080  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,414  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,244  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,064  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/233,063  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,397  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,399  
;; PRIOR FILING DATE: 2000-09-14

;; PRIOR APPLICATION NUMBER: 60/232,401  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,065  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,398

Query Match 38.7%; Score 41; DB 1; Length 45;  
Best Local Similarity 53.8%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GATFOVEYPSGSH 13  
Db 11 GAVIQTEISGEQH 23

RESULT 14  
US-09-764-874-5166  
;; Sequence 5166, Application US/09764874  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC010  
;; CURRENT FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 60/179,065  
;; PRIOR FILING DATE: 2000-01-31  
;; PRIOR APPLICATION NUMBER: 60/180,628  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: 60/214,886  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: 60/217,487  
;; PRIOR FILING DATE: 2000-07-11  
;; PRIOR APPLICATION NUMBER: 60/225,758  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/220,963  
;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: 60/217,496  
;; PRIOR FILING DATE: 2000-07-11  
;; PRIOR APPLICATION NUMBER: 60/225,447  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/218,290  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: 60/225,757  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/226,868  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/216,647  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,267  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/216,880  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,270  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/251,869  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/235,834  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/234,274  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: 60/234,223

PRIOR FILING DATE:	2000-09-21
PRIOR APPLICATION NUMBER:	60/228, 924A
PRIOR FILING DATE:	2000-08-30
PRIOR APPLICATION NUMBER:	60/224, 518
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/236, 365B
PRIOR FILING DATE:	2000-09-29
PRIOR APPLICATION NUMBER:	60/224, 519
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/220, 964A
PRIOR FILING DATE:	2000-07-26
PRIOR APPLICATION NUMBER:	60/241, 809D
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/249, 299D
PRIOR FILING DATE:	2000-11-01
PRIOR APPLICATION NUMBER:	60/225, 268B
PRIOR FILING DATE:	2000-08-14
PRIOR APPLICATION NUMBER:	60/236, 366B
PRIOR FILING DATE:	2000-09-29
PRIOR APPLICATION NUMBER:	60/224, 517
PRIOR FILING DATE:	2000-12-08
PRIOR APPLICATION NUMBER:	60/225, 866B
PRIOR FILING DATE:	2000-12-08
PRIOR APPLICATION NUMBER:	60/229, 344A
PRIOR FILING DATE:	2000-09-01
PRIOR APPLICATION NUMBER:	60/234, 997D
PRIOR FILING DATE:	2000-09-25
PRIOR APPLICATION NUMBER:	60/229, 343A
PRIOR FILING DATE:	2000-09-01
PRIOR APPLICATION NUMBER:	60/229, 345A
PRIOR FILING DATE:	2000-09-01
PRIOR APPLICATION NUMBER:	60/229, 287D
PRIOR FILING DATE:	2000-09-01
PRIOR APPLICATION NUMBER:	60/229, 513A
PRIOR FILING DATE:	2000-09-05
PRIOR APPLICATION NUMBER:	60/231, 413A
PRIOR FILING DATE:	2000-09-08
PRIOR APPLICATION NUMBER:	60/229, 509D
PRIOR FILING DATE:	2000-09-05
PRIOR APPLICATION NUMBER:	60/236, 367B
PRIOR FILING DATE:	2000-09-29
PRIOR APPLICATION NUMBER:	60/237, 039D
PRIOR FILING DATE:	2000-10-02
PRIOR APPLICATION NUMBER:	60/237, 038B
PRIOR FILING DATE:	2000-10-02
PRIOR APPLICATION NUMBER:	60/236, 370D
PRIOR FILING DATE:	2000-09-29
PRIOR APPLICATION NUMBER:	60/236, 802D
PRIOR FILING DATE:	2000-10-02
PRIOR APPLICATION NUMBER:	60/237, 037D
PRIOR FILING DATE:	2000-10-02
PRIOR APPLICATION NUMBER:	60/237, 040A
PRIOR FILING DATE:	2000-10-02
PRIOR APPLICATION NUMBER:	60/240, 966B
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/239, 935B
PRIOR FILING DATE:	2000-10-13
PRIOR APPLICATION NUMBER:	60/239, 937D
PRIOR FILING DATE:	2000-10-13
PRIOR APPLICATION NUMBER:	60/241, 787D
PRIOR FILING DATE:	2000-10-20
PRIOR APPLICATION NUMBER:	60/246, 474D
PRIOR FILING DATE:	2000-11-08
PRIOR APPLICATION NUMBER:	60/246, 552A
PRIOR FILING DATE:	2000-11-08
PRIOR APPLICATION NUMBER:	60/249, 216B
PRIOR FILING DATE:	2000-11-17

PRIOR APPLICATION NUMBER: 60/249, 210	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/226, 661	PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/225, 759	PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/225, 213	PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/227, 162	PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/225, 214	PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/235, 836	PRIOR FILING DATE: 2000-09-27	PRIOR APPLICATION NUMBER: 60/230, 438	PRIOR FILING DATE: 2000-09-06	PRIOR APPLICATION NUMBER: 60/215, 135	PRIOR FILING DATE: 2000-06-30	PRIOR APPLICATION NUMBER: 60/225, 266	PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/249, 218	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 208	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 213	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 212	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 207	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 245	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 244	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 217	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 211	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 215	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 264	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 214	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/249, 297	PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/232, 400	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/231, 242	PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/232, 081	PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/232, 080	PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/231, 414	PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/231, 244	PRIOR FILING DATE: 2000-09-08	PRIOR APPLICATION NUMBER: 60/233, 064	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/233, 063	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/232, 397	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/232, 399	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/232, 401	PRIOR FILING DATE: 2000-09-14	PRIOR APPLICATION NUMBER: 60/241, 808	PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/241, 826	PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/241, 786
---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------	-------------------------------	---------------------------------------

:  
 : PRIOR FILING DATE: 2000-10-20  
 : PRIOR APPLICATION NUMBER: 60/241,221  
 :  
 : PRIOR FILING DATE: 2000-10-20  
 : PRIOR APPLICATION NUMBER: 60/246,475  
 :  
 : PRIOR FILING DATE: 2000-11-08  
 : PRIOR APPLICATION NUMBER: 60/231,243  
 :  
 : PRIOR FILING DATE: 2000-09-08  
 : PRIOR APPLICATION NUMBER: 60/233,065  
 :  
 : PRIOR FILING DATE: 2000-09-14  
 : PRIOR APPLICATION NUMBER: 60/232,358

Query Match	38.7%	Score 41;	DB 21;	Length 45;
Best Local Similarity	53.8%;	Pred. No. 47;		
Matches	7;	Conservative	1;	Mismatches 5; Indels 0; Gaps 0;

```
QY      1 GATFOVEVPGSQH 13
          || | : | ||
Db      11 GAVLQTEISGEQH 23
```

```

RESULT 15
US-60-143-993-1325
: Sequence 1325, Application US/60143993
: GENERAL INFORMATION:
: APPLICANT: Kerlavage, Anthony
: TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
: TITLE OF INVENTION: THEREOF
: FILE REFERENCE: CLO00053
: CURRENT APPLICATION NUMBER: US/60/143,993
: CURRENT FILING DATE: 1993-07-15
: NUMBER OF SEQ ID NOS: 1906
: SOFTWARE: FASTSEQ for Windows Version 3.0.
: SEQ ID NO 1325
: LENGTH: 44
: TYPE: PRT
: ORGANISM: Drosophila
US-60-143-993-1325

```

Query Match	38.2%	Score 40.5;	DB 26;	Length 44;
Best Local Similarity	50.0%	Pred. No. 55;		
Matches 11; Conservative		3; Mismatches 7;	Indels 1;	Gaps 1;

```
QY 1 GATFQVEVP-GSQHIDSQKKAI 21
    ||::||| | | | |
Db 4 GADQRIEVPVGQVYDDQQLI 25
```

Search completed: October 24, 2002, 15:45:17  
Job time : 141.77 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:39 ; Search time 45.7869 Seconds  
(without alignments)  
137.008 Million cell updates/sec

Title: US-09-786-648-4

Sequence: 1 GATFOYVPGSGHDSOKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 378271

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Pending Patents\_AA\_New:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	46	43.4	99	US-10-217-607-761	Sequence 761, App
2	43	40.6	84	US-09-620-3938-3233	Sequence 3233, App
3	41	38.7	76	US-09-791-537-81337	Sequence 81337, A
4	39	36.8	60	US-10-219-999-51216	Sequence 51216, A
5	39	35.8	87	US-10-155-881-12211	Sequence 12211, A
6	38	35.8	51	US-10-182-995-28451	Sequence 28451, A
7	38	35.8	51	US-10-203-134-37782	Sequence 37782, A
8	38	35.8	51	US-10-203-136-37757	Sequence 37757, A
9	38	35.8	51	US-10-182-993-36805	Sequence 36805, A
10	38	35.8	51	US-10-182-997-26224	Sequence 26224, A
11	38	35.8	51	US-10-203-135-36160	Sequence 36160, A
12	38	35.8	51	US-10-203-137-37905	Sequence 37905, A
13	38	35.8	51	US-10-203-139-36373	Sequence 36373, A
14	38	35.8	76	US-09-791-537-137751	Sequence 137751, A
15	37	34.9	15	US-09-350-401A-3025	Sequence 3025, App
16	37	34.9	15	US-09-350-401A-3163	Sequence 3163, App
17	37	34.9	40	US-10-217-648-783	Sequence 783, App
18	37	34.9	45	US-09-791-537-151859	Sequence 151859, A
19	37	34.9	62	US-09-791-537-120702	Sequence 120702, A
20	37	34.9	62	US-09-791-537-120704	Sequence 120704, A
21	37	34.9	71	US-10-106-698-6660	Sequence 6660, App
22	37	34.9	75	US-10-232-643-3	Sequence 3, App11
23	37	34.9	76	US-09-791-537-81370	Sequence 81370, A
24	37	34.9	88	US-10-182-997-20295	Sequence 20295, A
25	37	34.9	88	US-10-182-998-11967	Sequence 11967, A
26	37	34.9	88	US-10-203-137-28226	Sequence 28226, A

27	37	34.9	88	6	US-10-203-138-12274	Sequence 12274, A
28	36	34.0	41	6	US-10-219-999-59647	Sequence 59647, A
29	36	34.0	44	6	US-10-182-998-14849	Sequence 14849, A
30	36	34.0	44	6	US-10-203-135-30068	Sequence 30068, A
31	36	34.0	44	6	US-10-203-138-15288	Sequence 15288, A
32	36	34.0	44	6	US-10-203-139-30215	Sequence 30215, A
33	36	34.0	48	5	US-09-791-537-74565	Sequence 74565, A
34	36	34.0	50	6	US-10-182-995-23617	Sequence 23617, A
35	36	34.0	50	6	US-10-182-995-27979	Sequence 27979, A
36	36	34.0	50	6	US-10-203-134-29975	Sequence 29975, A
37	36	34.0	50	6	US-10-203-134-37031	Sequence 37031, A
38	36	34.0	50	6	US-10-203-136-29998	Sequence 29998, A
39	36	34.0	50	6	US-10-203-136-37062	Sequence 37062, A
40	36	34.0	50	6	US-10-182-993-29373	Sequence 29373, A
41	36	34.0	50	6	US-10-182-993-36014	Sequence 36014, A
42	36	34.0	50	6	US-10-182-997-22308	Sequence 22308, A
43	36	34.0	50	6	US-10-182-997-25812	Sequence 25812, A
44	36	34.0	50	6	US-10-182-998-13901	Sequence 13901, A
45	36	34.0	50	6	US-10-203-135-35581	Sequence 35581, A

#### ALIGNMENTS

```
RESULT 1
US-10-217-607-761
Sequence 761, Application US/10217607
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PMO1C1N
CURRENT APPLICATION NUMBER: US/10/217,607
CURRENT FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: 09/735,463
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 761
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-217-607-761

Query Match          43.4%; Score 46; DB 6; Length 99;
Best Local Similarity 52.4%; Pred. No. 5.1;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY      1 GATFOYVPGSGHDSOKKAI 21
DB      74 GTTFVYGAGSGEDSEKRAI 94

RESULT 2
US-09-620-3938-3233
Sequence 3233, Application US/096203938
GENERAL INFORMATION:
APPLICANT: ALEXANDROV, Nikolai et al.
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
FILE REFERENCE: 2750-1068P
CURRENT APPLICATION NUMBER: US/09/620,3938
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 9948
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3233
```

```

; LENGTH: 84
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..84
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..84
; OTHER INFORMATION: Ceres Seq. ID 1385062
US-09-620-393B-3233

Query Match
Best Local Similarity 40.6%; Score 43; DB 5; Length 84;
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
Db 21 GKTYSMEGPGIQCDEHNKGL 41

RESULT 3
US-09-791-537-81337
; Sequence 81337, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81337
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Dictyostelium discoideum
US-09-791-537-81337

Query Match
Best Local Similarity 38.7%; Score 41; DB 5; Length 76;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21
Db 10 GKTTLEVGSDNIENYKAKI 30

RESULT 4
US-10-219-999-51216
; Sequence 51216, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovallig, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 51216
; LENGTH: 60
; TYPE: PRT
```

```

; ORGANISM: Zea mays
US-10-219-999-51216

Query Match
Best Local Similarity 36.8%; Score 39; DB 6; Length 60;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 TFOVEVPGSOHIDSOKKAI 21
Db 5 TLSVQVFGNHLVEFKKAI 23

RESULT 5
US-10-155-881-12211
; Sequence 12211, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovallig, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lufliyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 12211
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-12211

Query Match
Best Local Similarity 36.8%; Score 39; DB 6; Length 87;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 5 QVEVPGSOHIDSQ 17
Db 41 ELEVPAGAHIEDE 53

RESULT 6
US-10-182-995-28451
; Sequence 28451, Application US/10182995
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN HEART
; FILE REFERENCE: PB 0004 WO 1
; CURRENT APPLICATION NUMBER: US/10/182,995
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/652,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 29119
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 28451
```

```

; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109658.5
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AV651727.1, EVALU6 4.00e-15
US-10-182-995-28451
```

```

Query Match          35.8%; Score 38; DB 6; Length 51;
Best Local Similarity 46.2%; Pred. No. 56;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 7 EYVGSQHIDSOKK 19
      :|:| | :|:| |
Db 23 DLPKSHLSKKK 35
```

```

RESULT 7
US-10-203-134-37782
; Sequence 37782, Application US/10203134
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN BONE MARROW
; FILE REFERENCE: PB 0004 WO 6
; CURRENT APPLICATION NUMBER: US/10/203,134
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38628
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 37782
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109658.5
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AV651727.1, EVALU6 4.00e-15
US-10-203-134-37782
```

```

Query Match          35.8%; Score 38; DB 6; Length 51;
Best Local Similarity 46.2%; Pred. No. 56;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 7 EYVGSQHIDSOKK 19
      :|:| | :|:| |
Db 23 DLPKSHLSKKK 35
```

```

RESULT 8
US-10-203-136-37757
```

```

; Sequence 37757, Application US/10203136
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN ADULT LIVER
; FILE REFERENCE: PB 0004 WO 3
; CURRENT APPLICATION NUMBER: US/10/203,136
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38578
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 37757
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109658.5
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.3
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AV651727.1, EVALU6 4.00e-15
US-10-203-136-37757

Query Match          35.8%; Score 38; DB 6; Length 51;
Best Local Similarity 46.2%; Pred. No. 56;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 7 EYVGSQHIDSOKK 19
      :|:| | :|:| |
Db 23 DLPKSHLSKKK 35
```

```

RESULT 9
US-10-182-993-36805
; Sequence 36805, Application US/10182993
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN BRAIN
; FILE REFERENCE: PB 0004 WO 2
; CURRENT APPLICATION NUMBER: US/10/182,993
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
```

;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
;; NUMBER OF SEQ ID NOS: 37811  
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
;; SEQ ID NO 36805  
;; LENGTH: 51  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL109658.5  
;; FEATURE:  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4  
;; FILE REFERENCE: PB 0004 WO 5  
US-10-182-993-36805  
Query Match 35.8%; Score 38; DB 6; Length 51;  
Best Local Similarity 46.2%; Pred. No. 56;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 7 EYPSQHIDSOKK 19  
Db 23 DLPKSHLSKKK 35  
RESULT 10  
US-10-182-997-26224  
;; Sequence 26224, Application US/10182997  
;; GENERAL INFORMATION:  
;; APPLICANT: Molecular Dynamics, Inc.  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: PB 0004 WO 10  
;; CURRENT APPLICATION NUMBER: US/10/182,997  
;; PRIOR FILING DATE: 2002-08-02  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
;; NUMBER OF SEQ ID NOS: 26941  
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
;; SEQ ID NO 26224  
;; LENGTH: 51  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL109658.5  
;; FEATURE:  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4  
;; FILE REFERENCE: PB 0004 WO 12  
US-10-182-997-26224  
Query Match 35.8%; Score 38; DB 6; Length 51;  
Best Local Similarity 46.2%; Pred. No. 56;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 EYPSQHIDSOKK 19  
Db 23 DLPKSHLSKKK 35  
RESULT 11  
US-10-203-135-36160  
;; Sequence 36160, Application US/10203135  
;; GENERAL INFORMATION:  
;; APPLICANT: Molecular Dynamics, Inc.  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
;; FILE REFERENCE: PB 0004 WO 5  
;; CURRENT APPLICATION NUMBER: US/10/203,135  
;; CURRENT FILING DATE: 2002-08-02  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
;; NUMBER OF SEQ ID NOS: 37012  
;; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
;; SEQ ID NO 36160  
;; LENGTH: 51  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL109658.5  
;; FEATURE:  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6  
;; FILE REFERENCE: PB 0004 WO 7  
US-10-203-135-36160  
Query Match 35.8%; Score 38; DB 6; Length 51;  
Best Local Similarity 46.2%; Pred. No. 56;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 7 EYPSQHIDSOKK 19  
Db 23 DLPKSHLSKKK 35  
RESULT 12  
US-10-203-137-37905  
;; Sequence 37905, Application US/10203137  
;; GENERAL INFORMATION:  
;; APPLICANT: Molecular Dynamics, Inc.  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
;; FILE REFERENCE: PB 0004 WO 7  
;; CURRENT APPLICATION NUMBER: US/10/203,137  
;; CURRENT FILING DATE: 2002-08-02  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
;; PRIOR APPLICATION NUMBER: US 60/207,456



```

; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 37905
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109658.5
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AV651727.1, EVALUO 4.00e-15
US-10-203-137-37905

Query Match
Best Local Similarity 35.8%; Score 38; DB 6; Length 51;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 EVPGSOHDSOKK 19
Db 23 DLPKSHLSKKK 35

RESULT 13
US-10-203-139-36373
; Sequence 36373, Application US/10203139
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN FETAL LIVER
; FILE REFERENCE: PB 0004 WO 4
; CURRENT APPLICATION NUMBER: US/10/203,139
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 37156
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 36373
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109658.5
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
; FEATURE:
```

```

; OTHER INFORMATION: EST_HUMAN HIT: AV651727.1, EVALUO 4.00e-15
US-10-203-139-36373

Query Match
Best Local Similarity 35.8%; Score 38; DB 6; Length 51;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 EVPGSOHDSOKK 19
Db 23 DLPKSHLSKKK 35

RESULT 14
US-09-791-537-137751
; Sequence 137751, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 137751
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-09-791-537-137751

Query Match
Best Local Similarity 35.8%; Score 38; DB 5; Length 76;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEPGSOHDSOKKAI 21
Db 10 GKTTIEVSSDTIDNVKQKI 30

RESULT 15
US-09-350-401A-3025
; Sequence 3025, Application US/09350401A
; GENERAL INFORMATION:
; APPLICANT: Alessandro Sette
; APPLICANT: John Sidney
; APPLICANT: Scott Southwood
; APPLICANT: Maria A. Vitello
; APPLICANT: Brian D. Livingston
; APPLICANT: Esteban Celis
; APPLICANT: Ralph T. Kubo
; APPLICANT: Howard M. Grey
; APPLICANT: Robert Chesnut
; TITLE OF INVENTION: INDUCING CELLULAR IMMUNE RESPONSES TO
; TITLE OF INVENTION: HEPATITIS B VIRUS USING PEPTIDE AND NUCLEIC ACID
; FILE REFERENCE: 2060.0060008
; CURRENT APPLICATION NUMBER: US/09/350,401A
; CURRENT FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: US 08/820,360
; PRIOR FILING DATE: 1997-03-12
; PRIOR APPLICATION NUMBER: US 60/013,363
; PRIOR FILING DATE: 1996-03-13
; PRIOR APPLICATION NUMBER: US 09/189,702
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: US 08/205,713
; PRIOR FILING DATE: 1994-03-04
; PRIOR APPLICATION NUMBER: US 08/159,184
; PRIOR FILING DATE: 1993-11-29
; PRIOR APPLICATION NUMBER: US 08/073,205
; PRIOR FILING DATE: 1993-07-04
; NUMBER OF SEQ ID NOS: 3877
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3025
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially Synthesized Peptide
US-09-350-401A-3025

```

```

Query Match          34.9%; Score 37; DB 5; Length 15;
Best Local Similarity 61.5%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      3 TFOVEVPGSGHID 15
      : 1 1 1 1 1 1 1
Db      3 SFGVEPSPGSGHID 15

```

Search completed: October 24, 2002, 15:47:34  
Job time : 45.7869 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:31:03 : Search time 15.8361 Seconds  
(without alignments)  
127.423 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATFGYEVPGSGHDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 37652

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	44	41.5	91	2	C96580 hypothetical prote
2	41	38.7	91	2	F81109 conserved hypotnet
3	40	37.7	79	2	C95152 1566 family elemen
4	40	37.7	79	2	H95167 1566 family elemen
5	38	35.8	79	2	B98034 degenerate transpo
6	36	34.0	41	2	T36863 probable two-compo
7	36	34.0	76	1	U00A ubiquitin - oat
8	36	34.0	76	1	S12114 polyubiquitin - in
9	36	34.0	77	2	S19799 ubiquitin - potato
10	36	34.0	77	2	S16263 ubiquitin precursor
11	36	34.0	82	2	D75321 ribosomal protein
12	36	34.0	88	1	U05Y ubiquitin precursor
13	36	34.0	98	2	S28243 NADH dehydrogenase
14	36	34.0	99	2	S12161 polyubiquitin - la
15	36	34.0	99	2	S62680 ubiquitin/ribosoma
16	35	33.0	70	2	G81388 acyl carrier prote
17	34	32.1	70	2	T39058 hypothetical prote
18	34	32.1	80	2	A12538 hypothetical prote
19	34	32.1	80	2	F87520 hypothetical prote
20	33	31.1	70	2	D75413 hypothetical prote
21	33	31.1	95	2	T14743 hypothetical prote
22	32	30.2	35	2	T02997 sigma-E factor reg
23	32	30.2	38	2	F60657 circumsporozoite p
24	32	30.2	63	2	T43808 zinc finger protei
25	32	30.2	77	1	U0UT ubiquitin - trypan
26	32	30.2	77	1	S77465 acyl carrier prote
27	32	30.2	78	1	U0UTC ubiquitin precursor
28	32	30.2	78	1	D81025 H+-transporting AT
29	32	30.2	79	2	S77831 hypothetical prote

30	32	30.2	85	2	J01704 hypothetical 10K p
31	32	30.2	90	2	E71527 hypothetical prote
32	32	30.2	90	2	AE0277 probable phage-rel
33	32	30.2	91	2	H82370 conserved hypotnet
34	32	30.2	93	2	JC7711 SH3 domain binding
35	32	30.2	97	2	G72117 conserved hypotnet
36	32	30.2	97	2	B86506 C7031 hypothetical
37	32	30.2	98	1	F6FM ferredoxin [2Fe-2S
38	32	30.2	98	1	T17294 T-cell receptor be
39	32	30.2	99	1	S25233 ferredoxin [2Fe-2S
40	32	30.2	99	2	AE2324 ferredoxin I (limo
41	31.5	29.7	74	2	G83393 hypothetical prote
42	31.5	29.7	82	2	T45054 hypothetical prote
43	31.5	29.7	93	2	G64978 hypothetical 10.1
44	31.5	29.7	96	2	B59063 hypothetical prote
45	31.5	29.7	98	1	F6FM ferredoxin [2Fe-2S

#### ALIGNMENTS

```

RESULT 1
C96580
hypothetical protein F1511.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96580
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: C96580
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-91 <STO>
A:Cross-references: GB:AE005173; NID:q4587538; PIDN:AAU25769.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1511.6
A:Map position: 1
Query Match 41.5%; Score 44; DB 2; Length 91;
Best Local Similarity 47.6%; Pred. No. 5.3;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
QY 1 GATFGYEVPGSGHDSQKKAI 21
Db 58 GTFMLEVKESRIIOQVKMI 78
RESULT 2
F81109
conserved hypothetical protein NMB1218 [imported] - Neisseria meningitidis (strain MC
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
C:Accession: F81109; F81907
R:Rettelin, H.; Saunders, N.J.; Heideberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizsa, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moron, E.R.; Rappunli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: F81109
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-91 <TEF>

```

A:Cross-references: GB:AE002469; GB:AE002098; NID:g7226446; PIDN:AAFA1600.1; PID:g7226446  
A:Experimental source: serogroup B, strain MC58  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
N.; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: AB1775; MUID:20222536  
A:Accession: F81907  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <PAR>  
A:Cross-references: GB:AL152755; GB:AL157959; NID:g7379742; PIDN:CAB84626.1; PID:g738004  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMB1218; NMA1380

Query Match 38.7%; Score 41; DB 2; Length 91;  
Best Local Similarity 38.1%; Pred. No. 17;  
Matches 8; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
||| ||| : ||| :  
Db 61 GATFOVAVKVENQEDNDNYRAL 81

RESULT 3  
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: C95152  
R:Uetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A>Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: C95152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75412.1; PID:g14972794; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI314

Query Match 37.7%; Score 40; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEVPGSOHIDSOKK 19  
||| ||| : ||| :  
Db 26 VEIPQSRPFDSKKR 39

RESULT 4  
H95167  
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: H95167  
R:Uetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A>Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95167  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75537.1; PID:g14972931; GSPDB:GN00164; TIGR:  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI443

Query Match 37.7%; Score 40; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 21;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEVPGSOHIDSOKK 19  
||| ||| : ||| :  
Db 26 VEIPQSRPFDSKKR 39

RESULT 5  
B98034  
degenerate transposase (orf1) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: B98034  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A>Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: B98034  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <KUR>  
A:Cross-references: GB:AE007317; PIDN:ALU0103.1; PID:g15458942; GSPDB:GN00174  
C:Genetics:  
A:Gene: ISI381-truncation

Query Match 35.8%; Score 38; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 45;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEVPGSOHIDSOKK 19  
||| ||| : ||| :  
Db 26 VEIPQSRPFDSKKR 39

RESULT 6  
T36863  
probable two-component sensor kinase - Streptomyces coelicolor (fragment)  
C:Species: Streptomyces coelicolor  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36863  
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A  
submitted to the EMBL Data Library, June 1999  
A:Reference number: Z21616  
A:Accession: T36863  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-41 <SE>  
A:Cross-references: EMBL:AL079332; PIDN:CAB45294.1; GSPDB:GN00070; SCODB:SC15.10C  
C:Genetics:  
A:Gene: SCODB:SC15.10C

Query Match 34.0%; Score 36; DB 2; Length 41;  
Best Local Similarity 60.0%; Pred. No. 48;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATFOVEVPG 10  
||| ||| : ||| :  
Db 31 GTTTEMEVPG 40

RESULT 7  
U00A

ubiquitin - oat  
C:Species: Avena sativa (oat)  
C:Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 16-Aug-1996  
C:Accession: A02576  
R:Vierstra, R.D.; Langan, S.M.; Schaller, G.E.  
Biochemistry 25, 3105-3108, 1986  
A:Title: Complete amino acid sequence of ubiquitin from the higher plant Avena sativa.  
A:Reference number: A02576  
A:Accession: A02576  
A:Molecule type: Protein  
A:Residues: 1-76 <VIE>  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: protein degradation  
F:1-76/Domain: ubiquitin homology <UBH>  
F:76/Cross-link: Isopeptide carboxyl end (Gly) (interchain to Lys N6-amino of other prot

Query Match 34.0%; Score 36; DB 1; Length 76;  
Best Local Similarity 42.9%; Pred. No. 92;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKRAI 21  
| | : | | | : | | : | |  
DB 10 GKTTLEVESDPTIDNVKAKI 30

RESULT 8  
S12114  
polyubiquitin - inky cap (Coprinus congregatus) (fragment)  
C:Species: Coprinus congregatus  
C:Date: 21-Nov-1993 #sequence\_revision 17-Jul-1998 #text\_change 23-Jul-1999  
C:Accession: S12114  
R:Roeder, L.M.; Loftus, M.G.; Ross, I.K.  
Nucleic Acids Res. 18, 6449, 1990  
A:Title: A novel form of ubiquitin found in the basidiomycete fungus, Coprinus congregatus  
A:Reference number: S12114; MUID:91057153  
A:Accession: S12114  
A:Molecule type: DNA  
A:Residues: 1-76 <ROS>  
A:Cross-References: EMBL:X54672; NID:g2550; PIDD:CAA38483.1; PID:g2551  
A:Experimental source: strain californian, CC 16 homokaryon, mating type 1  
C:Function:  
A:Description: protein degradation  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: nucleus; polypeptide; protein degradation  
F:1-76/Product: ubiquitin #status predicted <UBI>  
F:1-76/Domain: ubiquitin homology <UBH>

Query Match 34.0%; Score 36; DB 2; Length 76;  
Best Local Similarity 42.9%; Pred. No. 92;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKRAI 21  
| | : | | | : | | : | |  
DB 10 GKTTLEVESDPTIDNVKAKI 30

RESULT 9  
S19799  
ubiquitin - potato  
C:Species: Solanum tuberosum (potato)  
C:Date: 30-Jun-1992 #sequence\_revision 03-Nov-2000 #text\_change 03-Nov-2000  
C:Accession: S19799; S19801; S19800  
R:Belknap, W.  
submitted to the EMBL Data Library, January 1992  
A:Reference number: S19798  
A:Accession: S19799  
A:Molecule type: mRNA  
A:Residues: 1-56 <BEL1>  
A:Cross-References: EMBL:211671; NID:g21603; PIDD:CAA77737.1; PID:g21604  
A:Accession: S19801  
A:Molecule type: mRNA  
A:Residues: 1-48 <BEL2>  
A:Cross-References: EMBL:211672; NID:g21607; PIDD:CAA77738.1; PID:g21608

A:Accession: S19800  
A:Molecule type: mRNA  
A:Residues: 37-77 <BEL3>  
A:Cross-References: EMBL:211673; NID:g21605; PIDD:CAA77739.1; PID:g21606  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: protein degradation  
F:1-76/Domain: ubiquitin homology <UBI>

Query Match 34.0%; Score 36; DB 2; Length 77;  
Best Local Similarity 42.9%; Pred. No. 93;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKRAI 21  
| | : | | | : | | : | |  
DB 10 GKTTLEVESDPTIDNVKAKI 30

RESULT 10  
S16263  
ubiquitin precursor - wheat (fragment)  
C:Species: Triticum aestivum (common wheat)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Jul-1999  
C:Accession: S16263  
R:Joshil, C.P.; Weng, J.; Nguyen, H.T.  
Plant Mol. Biol. 16, 907-908, 1991  
A:Title: Wheat ubiquitin gene exhibits a conserved protein coding region and a diverg  
A:Reference number: S16263; MUID:91316227  
A:Accession: S16263  
A:Molecule type: mRNA  
A:Residues: 1-77 <JOS>  
A:Cross-References: EMBL:X56601; NID:g21899; PIDD:CAA39938.1; PID:g21900  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: polypeptide  
F:1-76/Product: ubiquitin #status predicted <MAT>  
F:1-76/Domain: ubiquitin homology <UBH>

Query Match 34.0%; Score 36; DB 2; Length 77;  
Best Local Similarity 42.9%; Pred. No. 93;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKRAI 21  
| | : | | | : | | : | |  
DB 10 GKTTLEVESDPTIDNVKAKI 30

RESULT 11  
D75321  
ribosomal protein l33 - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 19-May-2000  
C:Accession: D75321  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: D75321  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-82 <WHI>  
A:Cross-References: GB:AE002041; GB:AE000513; NID:g6459838; PIDD:AAFI1599.1; PID:g645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2049  
A:Map position: 1  
C:Superfamily: Escherichia coli ribosomal protein l33

Query Match 34.0%; Score 36; DB 2; Length 82;  
Best Local Similarity 47.6%; Pred. No. 1e+02;  
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 GATFOVEVPGSQHIDSOKRAI 21

Db 9 GATP--PVPGDQREGCAKEST 27

## RESULT 12

UOSY  
ubiquitin precursor - soybean (fragment)  
N:Alternate names: polyubiquitin  
C:Species: Glycine max (soybean)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Aug-1996  
C:Accession: S02222  
R:Fortin, M.G.; Purohit, S.K.; Verma, D.P.S.  
Nucleic Acids Res. 16, 11377, 1988  
A:Title: The primary structure of soybean (glycine max) ubiquitin is identical to other  
A:Reference number: S02222; MUID:89083578  
A:Accession: S02222  
A:Molecule type: mRNA  
A:Residues: 1-88 <FOR>  
A:Cross-references: EMBL:X13251  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: polypeptide; protein degradation  
F:1-11/Product: ubiquitin (fragment) #status predicted <UBI1>  
F:12-87/Product: ubiquitin #status predicted <UBI2>  
F:12-87/Domain: ubiquitin homology <UBH>

Query Match 34.0%; Score 36; DB 1; Length 88;  
Best Local Similarity 42.9%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 21 GKTITLEVSSDTIDNVKAKI 41

## RESULT 13

S28243  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain C1-B12 - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 02-Dec-1993 #sequence\_revision 13-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S28243  
R:Walker, J.E.; Arimendi, J.M.; Dupuis, A.; Fearney, I.M.; Finel, M.; Medd, S.M.; Pili  
J. Mol. Biol. 226, 1051-1072, 1992  
A:Title: Sequences of 20 subunits of NADH: ubiquinone oxidoreductase from bovine heart  
A:Reference number: S28237; MUID:92389317  
A:Accession: S28243  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-98 <MWL>  
A:Cross-references: EMBL:X63213; NID:9235; PIDN:CAA44898.1; PID:9236  
C:Keywords: NAD; oxidoreductase

Query Match 34.0%; Score 36; DB 2; Length 98;  
Best Local Similarity 38.9%; Pred. No. 1.2e+02;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 AATFOVEVPGSQHDSOKK 19  
Db 76 AAFVVAIGAEVYLESSCKK 93

## RESULT 14

S12161  
polyubiquitin - large-leaved lupine (fragment)  
C:Species: Lupinus polyphyllus (large-leaved lupine)  
C:Date: 21-Nov-1993 #sequence\_revision 17-Jul-1998 #text\_change 23-Jul-1999  
C:Accession: S12161  
R:Perrey, R.; Warskulat, U.; Wink, M.  
Nucleic Acids Res. 18, 6428, 1990  
A:Title: Molecular cloning of a cDNA for the ubiquitin gene of Lupinus polyphyllus.  
A:Reference number: S12161; MUID:91057134  
A:Accession: S12161  
A:Molecule type: mRNA  
A:Residues: 1-99 <PER>

A:Cross-references: EMBL:X54381; NID:919513; PIDN:CAA38256.1; PID:919514  
A:Experimental source: leaf  
C:Function:

A:Description: protein degradation  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: nucleus; polypeptide; protein degradation  
F:1-22/Product: ubiquitin (fragment) #status predicted <UBI1>  
F:1-22/Domain: ubiquitin homology (fragment) #status predicted <UBH1>  
F:23-98/Product: ubiquitin #status predicted <UBI2>  
F:23-98/Domain: ubiquitin homology <UBH2>

Query Match 34.0%; Score 36; DB 2; Length 99;  
Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 32 GKTITLEVSSDTIDNVKAKI 52

## RESULT 15

S62680  
ubiquitin/ribosomal protein S27a fusion protein - Emeritella nidulans  
C:Species: Emeritella nidulans, Aspergillus nidulans  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
C:Accession: S62680  
R:Redkar, R.J.; Lemke, P.A.; Singh, N.K.  
Curr. Genet. 29, 130-135, 1996  
A:Title: Isolation of differentially expressed cDNA clones from salt-adapted Aspergillus  
A:Reference number: S62677; MUID:96418874  
A:Accession: S62680  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual  
A:Molecule type: mRNA  
A:Residues: 1-99 <RED>  
C:Superfamily: ubiquitin / rat ribosomal protein S27a; ribosomal protein S27a homolog  
F:1-76/Domain: ubiquitin homology <UBH>

Query Match 34.0%; Score 36; DB 2; Length 99;  
Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHDSOKKAI 21  
Db 10 GKTITLEVSSDTIDNVKTKI 30

Search completed: October 24, 2002, 15:37:45  
Job time : 17.8361 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:23:28 : Search time 7.57377 Seconds

(without alignments)  
107.359 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEYPSQSHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 12261

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : swissprot\_40:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	38.7	76	1	UBIQ_DICDI
2	38	35.8	76	1	P08618 dictyostell
3	37	34.9	76	1	P13117 neurospora
4	36	34.0	76	1	P04838 saccharomyc
5	36	34.0	76	1	P19848 coprinus co
6	36	34.0	76	1	P03993 glycine max
7	36	34.0	76	1	P02365 bos taurus
8	34	32.1	76	1	P49634 acanthamoeb
9	33	31.1	76	1	P22589 phytophthor
10	32	30.2	43	1	O91955 cyprinus ca
11	32	30.2	76	1	P42739 acetabulari
12	32	30.2	76	1	P42740 aglaetnam1
13	32	30.2	76	1	P14792 caenorhabdi
14	32	30.2	76	1	P14624 chlamydomon
15	32	30.2	76	1	P15174 trypanosoma
16	32	30.2	76	1	P08655 trypanosoma
17	32	30.2	77	1	P20804 synecocyst
18	32	30.2	88	1	O69250 bacillus me
19	32	30.2	97	1	O92560 chlamydia p
20	32	30.2	98	1	P06423 anabaena sp
21	32	30.2	98	1	FER1 ANASP
22	31.5	29.7	90	1	FER1 NOSMU
23	31.5	29.7	93	1	P22469 drosophila
24	31.5	29.7	93	1	YEHE_ECOLI
25	31	29.2	66	1	P33344 escherichia
26	31	29.2	66	1	P00248 mastigoclad
27	31	29.2	76	1	P58096 escherichia
28	31	29.2	76	1	P23324 euploties eu
29	31	29.2	80	1	P49633 leishmania
30	31	29.2	80	1	P23398 strongyloce
31	31	29.2	80	1	O50663 mycobacteri
32	31	29.2	91	1	P50715 mus musculu
33	31	29.2	91	1	O64263 mus musculu
			93	1	P17534 mus musculu
			93	1	P11777 mus musculu

34	31	29.2	93	1	FER2_EQUITE	P00236 equisetum t
35	30.5	28.8	17	1	PH3_PPRAM	P82696 periplaneta
36	30.5	28.8	93	1	FER2_EQUAR	P00237 equisetum a
37	30	28.3	39	1	YE04_HAEIN	P44179 haemophilus
38	30	28.3	53	1	YE80_ARCFU	O28792 archaeoglob
39	30	28.3	70	1	PYS2_FREDI	P14878 fremyella d
40	30	28.3	74	1	CSRA_BACSU	P33911 bacillus su
41	30	28.3	76	1	CSRA_BACSU	O92414 drosophila
42	30	28.3	76	1	UBIQ_HUMAN	P02248 homo sapien
43	30	28.3	76	1	UBIQ_LEITMA	O05550 leishmania
44	30	28.3	77	1	RS9_TOBAC	P49214 nicotiana t
45	30	28.3	90	1	YQAD_ECOLI	P76620 escherichia

#### ALIGNMENTS

RESULT 1  
UBIQ\_DICDI STANDARD: PRT: 76 AA.  
AC P08618;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1991 (Rel. 20, Last annotation update)  
DE Ubiquitin.  
OS Dictyostellum discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.  
OX NCBI\_Taxid=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88152253; PubMed=2831095;  
RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,  
RA Gerisch G.;  
RT "Complete cDNA sequence of a Dictyostellum ubiquitin with a carboxy-  
RT terminal tail and identification of the protein using an anti-peptide  
RT antibody.";  
RT FEBS Lett. 229:273-278(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89352609; PubMed=2548604;  
RA Ohmachi T., Giorda R., Shaw D.R., Ennis H.L.;  
RT "Molecular organization of developmentally regulated Dictyostellum  
RT discoideum ubiquitin cDNAs.";  
RN Biochemistry 28:5226-5231(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87257921; PubMed=3037345;  
RA Giorda R., Ennis H.L.;  
RT "Structure of two developmentally regulated Dictyostellum discoideum  
RT ubiquitin genes.";  
RN Mol. Cell. Biol. 7:2097-2103(1987).  
RN [4]  
RP SEQUENCE OF 13-76 FROM N.A.  
RC STRAIN=AX2;  
RA Westphal M., Mueller-Taubenberger A., Noegel A., Gerisch G.;  
RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
RT Dictyostellum discoideum.";  
RN FEBS Lett. 209:92-96(1986).  
RN [5]  
RP FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
CC REGULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
CC BIOGENESIS.  
CC -I- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
CC -I- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>)

CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).  
 -----  
 CC EMBL: X07210; CAA30183.1; ALT\_TERM.  
 DR EMBL: M19666; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19491; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19492; AAA33270.1; ALT\_TERM.  
 DR EMBL: M23748; AAA33262.1; ALT\_TERM.  
 DR EMBL: M23749; AAA33263.1; ALT\_TERM.  
 DR EMBL: M23750; AAA33265.1; ALT\_TERM.  
 DR EMBL: M23752; AAA33266.1; ALT\_TERM.  
 DR EMBL: M23753; AAA33267.1; ALT\_TERM.  
 DR EMBL: M23754; AAA33268.1; ALT\_TERM.  
 DR EMBL: X04702; CAA28408.1; ALT\_TERM.  
 DR PIR: S00357; UDOR.  
 DR PIR: E34080; UDOR7.  
 DR PIR: A27806; A27806.  
 DR PIR: B27806; B27806.  
 DR PIR: A34080; A34080.  
 DR PIR: B34080; B34080.  
 DR PIR: C34080; C34080.  
 DR PIR: D34080; D34080.  
 DR HSSP: P02248; 1UBI.  
 DR Dictydb: DD05001; ubqA.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; Ubiquitin.  
 DR SMART: SM00213; Ubq; 1.  
 DR PROSITE: PS00299; Ubiquitin\_1; 1.  
 DR PROSITE: PS50053; Ubiquitin\_2; 1.  
 DR Nucleic protein: Polyprotein.  
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76 76 MULTIBIQUITIN ADDUCTS.  
 FT VARIANT 11 11 CONJUGATION TO ACCEPTOR PROTEINS.  
 SQ SEQUENCE 76 AA; 8538 MW; 6427383968E8A84 CRC64;  
 K -> N (IN SOME CLONES REPEATS).  
 Query Match 38.7%; Score 41; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 5.3;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 OY 1 GATFOVEVPGSOHTDSOKKAI 21  
 DB 10 GKTTLEVESSDNTENYKAKI 30  
 -----  
 RESULT 2  
 ID UBIO\_NEUCR STANDARD; PRT; 76 AA.  
 AC P13117;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ubiquitin.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_Taxid=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;  
 RX MEDLINE=89366647; PubMed=2549509;  
 RA Taccoli G.E., Grotwold E., Aisenberg G.O., Judewicz N.D.;  
 RT "Ubiquitin expression in Neurospora crassa: cloning and sequencing of  
 a polyubiquitin gene.";  
 RL Nucleic Acids Res. 17:6153-6165(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST. LAWRENCE 74 / SL 74 / ORS 6A;  
 RX MEDLINE=91323720; PubMed=1650731;  
 RA Taccoli G.E., Grotwold E., Aisenberg G.O., Judewicz D.N.;  
 RT "The cDNA sequence and expression of an ubiquitin-tail gene fusion in  
 Neurospora crassa.";

RL Gene 102:133-137(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-74-OR23-1A;  
 RX MEDLINE=94374698; PubMed=8088539;  
 RA Tarawneh K.A., Annumla K.R., Free S.J.;  
 RT "The isolation and characterization of a Neurospora crassa gene  
 (ubi::crp-6) encoding a ubiquitin-40S ribosomal protein fusion  
 protein.";  
 RL Gene 147:137-140(1994).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: Ubiquitin IS SYNTHESIZED AS A POLYUBIQUITIN  
 PRECURSOR WITH 4 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL  
 AMINO-ACID (GIN) AFTER THE LAST REPEAT. SOME Ubiquitin GENES  
 CONTAIN A SINGLE COPY OF Ubiquitin FUSED TO A RIBOSOMAL PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC EMBL: X13140; CAA31530.1; ALT\_TERM.  
 DR EMBL: U01220; AAA56880.1; ALT\_TERM.  
 DR EMBL: U01221; AAA03351.1; ALT\_TERM.  
 DR PIR: S05323; UONC.  
 DR HSSP: P02248; 1UBI.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; Ubiquitin.  
 DR SMART: SM00213; Ubq; 1.  
 DR PROSITE: PS00299; Ubiquitin\_1; 1.  
 DR PROSITE: PS50053; Ubiquitin\_2; 1.  
 DR Nucleic protein: Polyprotein.  
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76 76 MULTIBIQUITIN ADDUCTS.  
 FT SEQUENCE 76 AA; 8598 MW; 994480EFD7D38403E CRC64;  
 SQ SEQUENCE 76 AA; 8598 MW; 994480EFD7D38403E CRC64;  
 K -> N (IN SOME CLONES REPEATS).  
 Query Match 35.8%; Score 38; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 17;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 OY 1 GATFOVEVPGSOHTDSOKKAI 21  
 DB 10 GKTTLEVESSDNTENYKAKI 30  
 -----  
 RESULT 3  
 ID UBIO\_YEAST STANDARD; PRT; 76 AA.  
 AC P04838;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Ubiquitin.  
 OS Candida albicans (Yeast), and  
 OS Cryptococcus neoformans (Filobasidiella neoformans).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_Taxid=4932, 5476, 5207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.cerevisiae;



RX MEDLINE=85061630; PubMed=6095120;  
 RA Oekavayak E., Finley D., Varshavsky A.;  
 RT "The yeast ubiquitin gene: head-to-tail repeats encoding a  
 RT polyubiquitin precursor protein.";  
 RL Nature 312:663-666(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (UBI1; UBI2; UBI3 AND UBI4).  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=87275838; PubMed=3038523;  
 RA Oekavayak E., Finley D., Solomon M.J., Varshavsky A.;  
 RT "The yeast ubiquitin genes: a family of natural gene fusions.";  
 RL EMBD J. 6:1429-1439(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A. (UBI2).  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=94205265; PubMed=8154186;  
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
 RT cerevisiae chromosome XI contains the UBI2 and MPI1 genes and three  
 RT new open reading frames.";  
 RL Yeast 9:1349-1354(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A. (UBI1).  
 RC SPECIES=S.cerevisiae; STRAIN=5288C / AB972;  
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (UBI3).  
 RC SPECIES=S.cerevisiae; STRAIN=5288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba D., Hallsworth K., Hawkins J., Hillier L., Jier T.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Ritten L., Riles L., Tatch A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A. (UBI4).  
 RC SPECIES=C.albicans;  
 RA Bailey D.A., Gow N.A.R., Brown A.J.P.;  
 RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A. (UBI1).  
 RC SPECIES=C.albicans; STRAIN=ATCC 26555;  
 RA Martinez J.P.;  
 RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C.neoformans; STRAIN=ATCC 6352;  
 RX MEDLINE=95369720; PubMed=7642124;  
 RA Splizer E.D., Splizer S.G.;  
 RT "Structure of the ubiquitin-encoding genes of Cryptococcus  
 RT neoformans.";  
 RL Gene 161:113-117(1995).  
 RN [9]  
 RP MUTAGENESIS OF LYS-29; LYS-48 AND GLY-76.  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=95340540; PubMed=7615550;  
 RA Johnson E.S., Ma P.C.M., Ota M., Varshavsky A.;  
 RT "A proteolytic pathway that recognizes ubiquitin as a degradation  
 RT signal.";  
 RL J. Biol. Chem. 270:17442-17456(1995).  
 RN [10]  
 RP MUTAGENESIS OF LYSINE RESIDUES.  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=95166210; PubMed=7862120;  
 FT

RA Spence J., Sadis S., Haas A.L., Finley D.;  
 RT "A ubiquitin mutant with specific defects in DNA repair and  
 RT multibubquitation.";  
 RL Mol. Cell. Biol. 15:1265-1273(1995).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS AND DNA REPAIR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 CC ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 CC POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 CC PROTEIN.  
 CC -1- MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 CC EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 CC AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 CC -1- MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 CC EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 CC AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X01473; CAA25704.1; ALT\_TERM.  
 DR EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 DR EMBL: X05728; CAA29195.1; ALT\_TERM.  
 DR EMBL: X05729; CAA29196.1; ALT\_TERM.  
 DR EMBL: X05730; CAA29197.1; ALT\_TERM.  
 DR EMBL: X05731; CAA29198.1; ALT\_TERM.  
 DR EMBL: X73541; CAA51949.1; ALT\_TERM.  
 DR EMBL: 228319; CAA82173.1; ALT\_TERM.  
 DR EMBL: 017246; AAB67466.1; ALT\_TERM.  
 DR EMBL: 254197; CAA90901.1; ALT\_TERM.  
 DR EMBL: 238059; CAA86130.1; ALT\_TERM.  
 DR EMBL: 032627; AAB4868.1; ALT\_TERM.  
 DR EMBL: 016992; AAB2979.1; ALT\_TERM.  
 DR PIR: A22696; UOBY.  
 DR PIR: C29456; UOBYR7.  
 DR PIR: A29456; A29456.  
 DR PIR: B29456; B29456.  
 DR PIR: D29456; D29456.  
 DR HSP: P02248; UBI1.  
 DR SWISS-2DPAGE: P04838; YEAST.  
 DR SGD: S0001410; RPL40A.  
 DR SGD: S0004157; RPS31.  
 DR SGD: S0003962; UBI4.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS00503; UBIQUITIN\_2; 1.  
 KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 FT -----  
 FT This SWISS-PROT entry is copyright. It is produced through a collaboration  
 FT between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 FT the European Bioinformatics Institute. There are no restrictions on its  
 FT use by non-profit institutions as long as its content is in no way  
 FT modified and this statement is not removed. Usage by and for commercial  
 FT entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 FT or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 FT -----  
 FT EMBL: X01473; CAA25704.1; ALT\_TERM.  
 FT EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 FT EMBL: X05728; CAA29195.1; ALT\_TERM.  
 FT EMBL: X05729; CAA29196.1; ALT\_TERM.  
 FT EMBL: X05730; CAA29197.1; ALT\_TERM.  
 FT EMBL: X05731; CAA29198.1; ALT\_TERM.  
 FT EMBL: X73541; CAA51949.1; ALT\_TERM.  
 FT EMBL: 228319; CAA82173.1; ALT\_TERM.  
 FT EMBL: 017246; AAB67466.1; ALT\_TERM.  
 FT EMBL: 254197; CAA90901.1; ALT\_TERM.  
 FT EMBL: 238059; CAA86130.1; ALT\_TERM.  
 FT EMBL: 032627; AAB4868.1; ALT\_TERM.  
 FT EMBL: 016992; AAB2979.1; ALT\_TERM.  
 FT PIR: A22696; UOBY.  
 FT PIR: C29456; UOBYR7.  
 FT PIR: A29456; A29456.  
 FT PIR: B29456; B29456.  
 FT PIR: D29456; D29456.  
 FT HSP: P02248; UBI1.  
 FT SWISS-2DPAGE: P04838; YEAST.  
 FT SGD: S0001410; RPL40A.  
 FT SGD: S0004157; RPS31.  
 FT SGD: S0003962; UBI4.  
 FT InterPro: IPR000626; Ubiquitin.  
 FT Pfam: PF00240; ubiquitin; 1.  
 FT PRINTS: PR00348; UBIQUITIN.  
 FT SMART: SM00213; UBO; 1.  
 FT PROSITE: PS00299; UBIQUITIN\_1; 1.  
 FT PROSITE: PS00503; UBIQUITIN\_2; 1.  
 FT KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION: INVOLVED IN THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 FT REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 FT BIOGENESIS AND DNA REPAIR.  
 FT SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 FT MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 FT ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 FT POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 FT PROTEIN.  
 FT MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUITIN. THERE IS A FINAL  
 FT AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 FT MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 FT EXACT HEAD TO TAIL REPEATS OF UBIQUIT

SQ SEQUENCE 76 AA; 8557 MW; 9AC936269D38403E CRC64;  
 Query Match 34.9%; Score 37; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 24;  
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSGHDSOKKAI 21  
 | | : | | : | | : | |  
 DB 10 GKTTTLEVESSTIDNVKSKI 30

RESULT 4  
 UBIO\_COPCO STANDARD; PRT; 76 AA.  
 AC P19848;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Ubiquitin.  
 OS Coprinus congregatus (Inky cap fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Coprinaceae; Coprinus.  
 ON NCBI\_TaxID=5347;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CALIFORNIA;  
 RX MEDLINE=91057153; PubMed=2173831;  
 RA Foster L.M., Loftus M.G., Ross I.K.;  
 RT "A novel form of ubiquitin found in the basidiomycete fungus,  
 RT Coprinus congregatus.";  
 RT Nucleic Acids Res. 18:6449-6449(1990).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54672; CAA38483.1; -  
 DR PIR; S12114; S12114.  
 DR HSSP; P02248; IUBI.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin.1.  
 DR PRINTS; PR00346; UBIOUITIN.  
 DR SMART; SM00213; UBO; 1.  
 DR PROSITE; PS00299; UBIOUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIOUITIN\_2; 1.  
 KW Nuclear protein; Polyprotein; Multigene family.  
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76 76 MULTIBIQUITIN ADDUCTS.  
 FT SEQUENCE 76 AA; 8541 MW; 852936397D38403E CRC64;  
 SQ SEQUENCE 76 AA; 8541 MW; 852936397D38403E CRC64;  
 Query Match 34.0%; Score 36; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 35;  
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSGHDSOKKAI 21  
 | | : | | : | | : | |  
 DB 10 GKTTTLEVESSTIDNVKSKI 30

RESULT 5  
 UBIO\_SOYBN STANDARD; PRT; 76 AA.  
 ID UBIO\_SOYBN  
 AC P03993;

DT 23-OCT-1986 (Rel. 02, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Ubiquitin.  
 OS Glycine max (Soybean).  
 OS Pisum sativum (Garden pea).  
 OS Avena sativa (Oat).  
 OS Avena fatua,  
 OS Hordeum vulgare (Barley),  
 OS Triticum aestivum (Wheat),  
 OS Zea mays (Maize),  
 OS Helianthus annuus (Common sunflower),  
 OS Arabidopsis thaliana (Mouse-ear cress),  
 OS Lupinus albus (White lupine),  
 OS Lupinus polyphyllus (Large-leaved lupine),  
 OS Lycopersicon esculentum (Tomato),  
 OS Solanum tuberosum (Potato),  
 OS Linum usitatissimum (Flax) (linseed),  
 OS Nicotiana glauca (Wood tobacco),  
 OS Asparagus officinalis (Garden asparagus),  
 OS Oryza sativa (Rice),  
 OS Petroselinum crispum (Parsley) (Petroselinum hortense), and  
 OS Brassica rapa (Turnip).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Glycine.  
 ON NCBI\_TaxID=3847, 3888, 4498, 4499, 4513, 4565, 4577, 4232, 3702, 3870,  
 OX 3874, 4081, 4113, 4006, 4096, 4686, 4530, 4043, 51350;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=soybean;  
 RX MEDLINE=89083578; PubMed=2849766;  
 RA Fortin M.G., Purohit S.K., Verma D.P.S.;  
 RT "The primary structure of soybean (Glycine max) ubiquitin is  
 RT identical to other plant ubiquitins.";  
 RT Nucleic Acids Res. 16:11377-11377(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=soybean;  
 RX MEDLINE=94211917; PubMed=8159798;  
 RA Bao-Gen X., Waterhouse R.N., Watanabe Y., Kajiwara H., Komatsu S.,  
 RA Hirano H.;  
 RT "Nucleotide sequence of a soybean (Glycine max L. Merr.) ubiquitin  
 RT gene.";  
 RT Plant Physiol. 104:805-806(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P. sativum;  
 RA Watts F.Z., Moore A.L.;  
 RT "Nucleotide sequence of a full length cDNA clone encoding a  
 RT polyubiquitin gene from Pisum sativum.";  
 RT Nucleic Acids Res. 17:10100-10100(1989).  
 RN [4]  
 RP SEQUENCE.  
 RC SPECIES=A. sativa;  
 RA Vierstra R.D., Langan S.M., Schaller G.E.;  
 RT "Complete amino acid sequence of ubiquitin from the higher plant  
 RT Avena sativa.";  
 RT Biochemistry 25:3105-3108(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (UBQ4).  
 RC SPECIES=A. thaliana; STRAIN=CV. COLUMBIA;  
 RX MEDLINE=89039731; PubMed=2460733;  
 RA Burke T.J., Callis J., Vierstra R.D.;  
 RT "Characterization of a polyubiquitin gene from Arabidopsis thaliana.";  
 RT Mol. Gen. Genet. 213:435-443(1988).  
 RN [6]  
 RP SEQUENCE FROM N.A. (UBQ1; UBQ2; UBQ5 AND UBQ6).  
 RC SPECIES=A. thaliana; STRAIN=CV. COLUMBIA;  
 RX MEDLINE=90324239; PubMed=2165066;  
 RA Callis J., Raesch J.A., Vierstra R.D.;  
 RT "Ubiquitin extension proteins of Arabidopsis thaliana. Structure,  
 RT localization, and expression of their promoters in transgenic

RT tobacco. ";  
 RL J. Biol. Chem. 265:12486-12493(1990).  
 RN [7]  
 RP SEQUENCE FROM N.A. (UBQ13).  
 RC SPECIES=A.thaliana;  
 RX MEDLINE=93177218; PubMed=8382544;  
 RA Sun C.W., Callis J.;  
 RL "Recent stable insertion of mitochondrial DNA into an Arabidopsis  
 polyubiquitin gene by nonhomologous recombination.";  
 RT Plant Cell 5:97-107(1993).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=H.vulgaris; STRAIN=CV. BOMI;  
 RX MEDLINE=86274685; PubMed=2426105;  
 RA Gausing K., Barkardottir R.;  
 RL "Structure and expression of ubiquitin genes in higher plants.";  
 RT Eur. J. Biochem. 158:57-62(1986).  
 RN [9]  
 RP SEQUENCE FROM N.A. (MUB1 AND MUB2).  
 RC SPECIES=H.vulgaris; STRAIN=CV. BOMI;  
 RX MEDLINE=91078635; PubMed=1701748;  
 RA Gausing K., Jensen C.B.;  
 RL "Two ubiquitin-long-tail fusion genes arranged as closely spaced  
 direct repeats in barley.";  
 RT Gene 94:165-171(1990).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=T.aestivum; STRAIN=CV. MUSTANG; TISSUE=Leaf;  
 RX MEDLINE=91316227; PubMed=1650258;  
 RA Joshi C.P., Nguyen H.T.;  
 RL "Wheat ubiquitin gene exhibits a conserved protein coding region and  
 a diverged 3' non-coding region.";  
 RT Plant Mol. Biol. 16:907-908(1991).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=H. annuus; STRAIN=CV. HA401B / CARGILL;  
 RX MEDLINE=89183614; PubMed=2538802;  
 RA Bluet M.N., Steilmetz A., Tessier L.H.;  
 RL "The primary structure of sunflower (Helianthus annuus) ubiquitin.";  
 RT Nucleic Acids Res. 17:2119-2119(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Maize;  
 RX Christensen A.H., Quail P.H.;  
 RL "Sequence analysis and transcriptional regulation by heat shock of  
 polyubiquitin transcripts from maize.";  
 RT Plant Mol. Biol. 12:619-632(1989).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Maize;  
 RX MEDLINE=92084112; PubMed=1660830;  
 RA Chen K., Rubenstein I.;  
 RL "Characterization of the structure and transcription of an ubiquitin  
 fusion gene from maize.";  
 RT Gene 107:205-212(1991).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=L.polyphyllus;  
 RX MEDLINE=91057134; PubMed=2173830;  
 RA Berry R., Watskulat U., Wink M.;  
 RL "Molecular Cloning of a cDNA for the ubiquitin gene of Lupinus  
 polyphyllus.";  
 RT Nucleic Acids Res. 18:6428-6428(1990).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=L.albus;  
 RX MEDLINE=94156199; PubMed=8112604;  
 RA Jacinto A., Neves A.M., Vassilevska T.D., Ricardo C.P.,  
 RL Rodrigues-Pousada C.;  
 RT "Cloning and characterization of two ubiquitin-79-amino-acid  
 extension protein-encoding fusion genes from Lupinus albus.";  
 RT Gene 139:201-205(1994).  
 RN [16]

Query Match	Best Local Similarity	Score 36;	DB 1;	Length 76;
Matches 9;	Conservative 2;	Mismatches 10;	Indels 0;	Gaps 0;
QY	1	GATGQVEVPGSQHIDSOKKAI	21	
DB	10	GKTTLEVESSDTIDNVRKAI	30	
<p>RESULT 6</p> <p>NR2M_BOVIN STANDARD: PRT: 97 AA.</p> <p>AC 002365;</p> <p>DT 01-JUL-1993 (Rel. 26, Created)</p> <p>DT 01-JUL-1993 (Rel. 26, Last sequence update)</p> <p>DE 15-DEC-1998 (Rel. 37, Last annotation update)</p> <p>DE MADH-ubiquitinone oxidoreductase B12 subunit (EC 1.6.5.3) (EC 1.6.99.3)</p> <p>DE (Complex 1-B12) (Cl-B12).</p> <p>GN NDUFB3.</p> <p>Bos taurus (Bovine).</p>				

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 CC NCBI\_Taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX TISSUE=Heart; PubMed=1518044;  
 RX MEDLINE=92389317; PubMed=1518044;  
 RA Walker J.E., Arizumi J.M., Dupuis A., Fearney I.M., Finel M.,  
 RA Med S.M., Pilkington S.J., Runswick M.J., Skehel J.M.;  
 RT "Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from  
 RT bovine heart mitochondria. Application of a novel strategy for  
 RT sequencing proteins using the polymerase chain reaction.";  
 RN J. Mol. Biol. 226:1051-1072(1992).  
 RN [2]  
 RP ACETYLATION.  
 RA Fearney I.M.;  
 RL Unpublished observations (DEC-1998).  
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY  
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED  
 CC TO BE UBIQUINONE.  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X63213; CA444898.1; -.  
 DR PIR; S28243; S28243.  
 KM Oxidoreductase; NAD: Ubiquinone; Mitochondrion; Acetylation.  
 FT INIT.MET 0  
 FT MOD\_RES 1 1 ACETYLATION (PARTIAL).  
 SQ SEQUENCE 97 AA: 11009 MW: 6845E45A2B6FEF2E CRC64:  
 Query Match 34.0%; Score 36; DB 1; Length 97;  
 Best Local Similarity 38.9%; Pred. No. 46;  
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Oy 2 ATFOVEVPGSGHDSOKK 19  
 Db 75 AAFVAIGAEEYLESOKK 92  
 RESULT 7  
 UBIQ\_ACACA STANDARD; PRT; 76 AA.  
 ID UBIQ\_ACACA  
 AC P49634;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Ubiqlitin.  
 OS Acanthamoeba castellanii (Acanthamoeba).  
 CC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 CC NCBI\_Taxid=5755;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94250685; PubMed=8193157;  
 RA "An K.S., Henney H.R.;  
 RA "An Acanthamoeba ubiqlitin-fusion protein; cDNA and deduced protein  
 RT sequence.";  
 RL Biochim. Biophys. Acta 1218:109-111(1994).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC REGULATOR OF GENE EXPRESSION, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED WITH RIBOSOMAL CEP52 AS

CC ITS C-TERMINAL EXTENSION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X75628; CA53293.1; ALT\_TERM.  
 DR HSSP; P02248; UBI.  
 DR InterPro; IPR000626; Ubiqlitin.  
 DR Pfam; PF00240; ubiqlitin; 1.  
 DR PRINTS; PR00348; UBIQUITIN.  
 DR SMART; SM00213; UBO; 1.  
 DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 KM Nuclear protein; Polypeptide.  
 FT SITE 48  
 FT NUCLEAR PROTEIN. NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76  
 FT CONJUGATION TO ACCEPTOR PROTEINS.  
 SQ SEQUENCE 76 AA: 8596 MW: D84480E07D1E9E52 CRC64:  
 Query Match 32.1%; Score 34; DB 1; Length 76;  
 Best Local Similarity 38.1%; Pred. No. 75;  
 Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
 Oy 1 GATFOVEVPGSGHDSOKKAI 21  
 Db 10 GKTITLEVSSDTIENVAOKI 30  
 RESULT 8  
 UBIQ\_PHYIN STANDARD; PRT; 76 AA.  
 ID UBIQ\_PHYIN  
 AC P22589;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE Ubiqlitin.  
 OS Phytophthora infestans (Potato late blight fungus).  
 CC Eukaryota; Stramenopiles; Oomycetes; Fythiales; Fythyaceae;  
 CC Phytophthora.  
 CC NCBI\_Taxid=4787;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-ISOLATE 88069;  
 RX MEDLINE=92003691; PubMed=1655113;  
 RA Pieterse C.M.J., Risseuw E.P., Davidse L.C.;  
 RA "An in planta induced gene of Phytophthora infestans codes for  
 RT ubiqlitin.";  
 RL Plant Mol. Biol. 17:799-811(1991).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: THIS UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH 3 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL  
 CC AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X55717; CA439250.1; ALT\_TERM.  
 DR PIR; S17740; UOINT.  
 DR HSSP; P02248; UBT.

```

DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR PRINTS; PR00348; UBIQUITIN.
DR SMART; SM00213; UBO; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
DR NCBI; P00001; Ubiquitin.
FT SITE 48 48
FT NUCLEAR PROTEIN; POLYPROTEIN.
FT BINDING 76 76
FT SEQUENCE 76 AA; 8580 MW; 96F084FE7F3A456F CRC64;
SQ
Query Match 32.1%; Score 34; DB 1; Length 76;
Best Local Similarity 38.1%; Pred. No. 75;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSHIDSOKRAI 21
DB 10 GKTITLDEPSPSIDNVKOKI 30

RESULT 9
TIBA_CYPCA STANDARD; PRT; 45 AA.
AC Q91955;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Thymosin beta-a.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprinus.
CC NCBI_TaxId=7962;
RN [1]
RP SEQUENCE FROM N.A.
RA Fujiki K., Nakao M., Shin D., Yano T.;
RT "Molecular cloning of carp (Cyprinus carpio) thymosin beta a.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays an important role in the organization of the
CC cytoskeleton. Binds to and sequesters actin monomers (G actin) and
CC therefore inhibits actin polymerization (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE THYMOSIN BETA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB028456; BAA96492.1; -
CC InterPro; IPR001152; Thymosin_b4.
CC Pfam; PF01290; Thymosin; 1.
CC ProDom; PD005116; Thymosin_b4; 1.
CC SMART; SM00152; THY; 1.
CC PROSITE; PS00500; THYMOSIN_B4; 1.
CC Actin-binding; Cytoskeleton.
CC INIT_MET 0 0
CC SEQUENCE 45 AA; 5063 MW; D05BA129A422EB70 CRC64;
SQ
Query Match 31.1%; Score 33; DB 1; Length 45;
Best Local Similarity 46.2%; Pred. No. 62;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 VPGSQHIDSOKRA 20
DB 28 LPTKEDIDOKRA 40

RESULT 10

```

```

TYBO_TORMA STANDARD; PRT; 43 AA.
ID TYBO_TORMA
AC Q91960;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Thymosin beta-10 (Thyb10).
OS Torpedo marmorata (Marbled electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hypnosqualia; Pristigaster; Batoidae;
OC Torpediniformes; Torpedinidae; Torpedo.
CC NCBI_TaxId=7788;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-electric lobe;
RC O'Regan S., Matz V., Cha N., Meunier F.M.;
RT "Torpedo electric lobe cDNAs that suppress a choline metabolism
RT mutation in yeast.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays an important role in the organization of the
CC cytoskeleton. Binds to and sequesters actin monomers (G actin) and
CC therefore inhibits actin polymerization (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE THYMOSIN BETA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ276369; CAB76965.1; -
CC InterPro; IPR001152; Thymosin_b4.
CC Pfam; PF01290; Thymosin; 1.
CC ProDom; PD005116; Thymosin_b4; 1.
CC SMART; SM00152; THY; 1.
CC PROSITE; PS00500; THYMOSIN_B4; 1.
CC Actin-binding; Cytoskeleton.
CC INIT_MET 0 0
CC SEQUENCE 43 AA; 4785 MW; 63D5D49977A2B351 CRC64;
SQ
Query Match 30.2%; Score 32; DB 1; Length 43;
Best Local Similarity 46.2%; Pred. No. 86;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 VPGSQHIDSOKRA 20
DB 28 LPTKEDIDOKRA 40

RESULT 11
UBIQ_ACECL STANDARD; PRT; 76 AA.
ID UBIQ_ACECL
AC P42739;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Ubiquitin.
OS Acetabularia cliftonii.
OC Eukaryota; Viridiplantae; Chlorophyta; Ulvophyceae; Dasycladales;
OC Dasycladaceae; Acetabularia.
CC NCBI_TaxId=3137;
RN [1]
RP SEQUENCE FROM N.A.
RA Frank S., Menzel D.;
RT Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

```

CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. THERE ARE TWO  
 CC ADDITIONAL AMINO-ACIDS (ALA-PHE) AFTER THE LAST REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z28649; CA82268.1; ALT\_SEQ.  
 DR HSSP: P02248; IUBI.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 DR Nucleic protein; Polypeptide.  
 KW SITE 48  
 FT BINDING 76 76 NECESSARY FOR BRANCHED-CHAIN  
 FT VARIANT 18 18 MULTIBIQUITIN ADDUCTS.  
 FT VARIANT 54 54 CONJUGATION TO ACCEPTOR PROTEINS.  
 FT SEQUENCE 76 AA; 8541 MW; 34D636389D0903F CRC64;  
 SO  
 Query Match 30.2%; Score 32; DB 1; Length 76;  
 Best Local Similarity 33.3%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 GATFOVEYPSQHIDSOKKAI 21  
 Db 10 GKTITLEVSSDTIENVKSI 30  
 RESULT 12  
 ID UBIQ\_AGLNE STANDARD; PRT; 76 AA.  
 AC P42740;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Ubiquitin.  
 OS Aglaetothamnion neglectum.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales;  
 OC Aglaetothamnion.  
 OC NCBI\_TaxID=2765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Apt R.E., Grossman A.R.;  
 RT "A polyubiquitin cDNA from a red alga.";  
 RL Plant Physiol. 99:1732-1733(1992).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH SIX HEAD TO TAIL REPEATS. THERE IS ONE ADDITIONAL  
 CC AMINO-ACID (MET) AFTER THE LAST REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M88684; AAA72126.1; ALT\_TERM.  
 DR HSSP: P02248; IUBI.

DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 KW Nucleic protein; Polypeptide.  
 FT SITE 48  
 FT BINDING 76 76 NECESSARY FOR BRANCHED-CHAIN  
 FT VARIANT 8 8 MULTIBIQUITIN ADDUCTS.  
 FT VARIANT 61 61 CONJUGATION TO ACCEPTOR PROTEINS.  
 FT VARIANT 61 61 L -> I (IN ONE OF THE REPEATS).  
 FT VARIANT 61 61 I -> N (IN ONE OF THE REPEATS).  
 FT SEQUENCE 76 AA; 8585 MW; DE993623CCE9B52 CRC64;  
 SO  
 Query Match 30.2%; Score 32; DB 1; Length 76;  
 Best Local Similarity 38.1%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 GATFOVEYPSQHIDSOKKAI 21  
 Db 10 GKTITLEVSSDTIENVKSI 30  
 RESULT 13  
 ID UBIQ\_CAEEL STANDARD; PRT; 76 AA.  
 AC P14792;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Ubiquitin.  
 GN (UBQ-1 OR UBI1 OR F25B5.4) AND (UBQ-2 OR UBI2).  
 OS Caenorhabditis elegans, and  
 OS Geodia cydonium (Sponge).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239, 6047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=C.elegans;  
 RC MEDLINE=89181579; PubMed=2538720;  
 RA Graham R.W., Jones D., Candido E.P.M.;  
 RT "UbiA, the major polyubiquitin locus in Caenorhabditis elegans, has  
 RT unusual structural features and is constitutively expressed.";  
 RL Mol. Cell. Biol. 9:268-277(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=C.elegans; STRAIN=BRISTOL N2;  
 RA Tatch A., Waterston R.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-37 FROM N.A.  
 RP SPECIES=C.elegans;  
 RC MEDLINE=88273146; PubMed=2839490;  
 RA Graham R.W., Van Doren K., Bektesh S., Candido E.P.M.;  
 RT "Maturation of the major ubiquitin gene transcript in Caenorhabditis  
 RT elegans involves the acquisition of a trans-spliced leader.";  
 RL J. Biol. Chem. 263:10415-10419(1988).  
 RN [4]  
 RP SEQUENCE FROM N.A. (UBQ-2).  
 RP SPECIES=C.elegans; STRAIN=BRISTOL N2;  
 RC MEDLINE=96018742; PubMed=7556908;  
 RA Jones D., Candido P.M.;  
 RT "A portable regulatory element directs specific expression of the  
 RT Caenorhabditis elegans ubiquitin gene ubq-2 in the somatic gonad.";  
 RL Dev. Biol. 171:60-72(1995).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RP SPECIES=G.cydonium;  
 RC MEDLINE=94110294; PubMed=8282761;  
 RA Pfeiffer K., Frank K., Schroeder H.C., Gamulin V., Rinkevich B.,  
 RA Batel R., Mueller I.M., Mueller E.G.;

```

RT "Cloning of the polyubiquitin cDNA from the marine sponge Geodia
RT cydonium and its preferential expression during reaggregation of
RT cells";
RL J. Cell Sci. 106:545-554(1993).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH 11 EXACT HEAD TO TAIL REPEATS. THERE ARE TWO
CC ADDITIONAL AMINO-ACIDS (ASP-ILE) AFTER THE LAST REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M23433; AAA28154.1; ALT_TERM.
CC EMBL: U23172; AAC6525.1; ALT_TERM.
CC EMBL: M21321; AAA28153.1;
CC EMBL: L31492; AAC37252.1; ALT_TERM.
CC EMBL: X70917; CAA50268.1; ALT_TERM.
CC PIR: A30126; A30126.
CC PIR: A28804; A28804.
CC HSSP: P02248; 1UBI.
CC WormPep: F25B5.4; CE01921.
CC InterPro: IPR000626; Ubiquitin.
CC Pfam: PF00240; ubiquitin; 1.
CC PRINTS: PR00348; UBIQUITIN.
CC SMART: SM00213; UBO; 1.
CC PROSITE: PS00299; UBIQUITIN_1; 1.
CC PROSITE: PS50053; UBIQUITIN_2; 1.
CC KMW Nuclear protein; Polyprotein.
FT SITE 48 48
FT NUCLEOTIDE SEQUENCE: Polyprotein.
FT BINDING 76 76
FT CONJUGATION TO ACCEPTOR PROTEINS.
SQ SEQUENCE 76 AA; 8539 MW; C43B24396EC9B52 CRC64;

Query Match 30.2%; Score 32; DB 1; Length 76;
Best Local Similarity 38.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GATFOYEVPGSOHIDSQKAI 21
Db 10 GKTTITLEVSSDTTENVKAKI 30

RESULT 14
ID UBIQ_CHLRE STANDARD; PRT; 76 AA.
AC P14624;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Ubiquitin.
GN UBI1 AND UBI3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A. (UBI1).
RC STRAIN=2137;
RX MEDLINE=90045969; PubMed=2554258;
RA Callis J., Pollmann L., Shanklin J., Wettern M., Vierstra R.;
RT "Sequence of a cDNA from Chlamydomonas reinhardtii encoding a
RT ubiquitin 52 amino acid extension protein.";
RL Nucleic Acids Res. 17:8377-8377(1989).
RN [2]

```

```

RP SEQUENCE FROM N.A. (UBI3).
RX MEDLINE=92037644; PubMed=1657605;
RA Pollmann L., von Kampen J., Wettern M.;
RT Ubiquitin in a lower plant. Characterization of ubiquitin-encoding
RT DNA and RNA from Chlamydomonas reinhardtii.
RL Eur. J. Biochem. 202:197-204(1991).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: UBI1 IS SYNTHESIZED AS A POLYPROTEIN WITH ONE COPY
CC OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15427; CAA33466.1; ALT_TERM.
CC EMBL: X60826; CAA43216.1; ALT_TERM.
CC PIR: S06598; UOKM.
CC HSSP: P02248; 1UBI.
CC InterPro: IPR000626; Ubiquitin.
CC Pfam: PF00240; ubiquitin; 1.
CC PRINTS: PR00348; UBIQUITIN.
CC SMART: SM00213; UBO; 1.
CC PROSITE: PS00299; UBIQUITIN_1; 1.
CC PROSITE: PS50053; UBIQUITIN_2; 1.
CC KMW Nuclear protein; Polyprotein.
FT SITE 48 48
FT NUCLEOTIDE SEQUENCE: Polyprotein.
FT BINDING 76 76
FT CONJUGATION TO ACCEPTOR PROTEINS.
SQ SEQUENCE 76 AA; 8539 MW; C42936277D1EB52 CRC64;

Query Match 30.2%; Score 32; DB 1; Length 76;
Best Local Similarity 38.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GATFOYEVPGSOHIDSQKAI 21
Db 10 GKTTITLEVSSDTTENVKAKI 30

RESULT 15
ID UBIQ_TRYBB STANDARD; PRT; 76 AA.
AC P15174;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Ubiquitin.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90136711; PubMed=2559328;
RA Wong S., Campbell D.A.;
RT "A polyubiquitin gene from Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 37:147-150(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=91088342; PubMed=2175891;
RA Wong S., Morales T.H., Campbell D.A.;
RT "Ubiquitin-Ep52 fusion protein homologs from Trypanosoma brucei.";
RL Nucleic Acids Res. 18:7181-7181(1990).
RN [3]
RP SEQUENCE FROM N.A.

```

RC STRAIN=427;  
RX MEDLINE=93109304; PubMed=8380221;  
RA Wong S., Morales T.H., Neigel J.E., Campbell D.A.;  
RT "Genomic and transcriptional linkage of the genes for calmodulin, EF-  
RT hand 5 protein, and ubiquitin extension protein 52 in Trypanosoma  
RT brucei.";  
RL Mol. Cell. Biol. 13:207-216(1993).  
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
CC BIOGENESIS.  
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
CC PRECURSOR WITH 5 OR 6 EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN  
CC GENES CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL  
CC PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; X14554; CAA32691.1; ALT\_TERM.  
DR EMBL; X54641; CAA38453.1; -  
DR EMBL; X54642; CAA38454.1; -  
DR EMBL; X56511; CAA39863.1; -  
DR EMBL; X56511; CAA39864.1; -  
DR PIR; S07998; UOQT.  
DR PIR; S14197; S14197.  
DR HSSP; P02248; IOBI.  
DR InterPro; IPR000626; Ubiqultin.  
DR Pfam; PF00240; ubiultin; 1.  
DR PRINTS; PR00348; UBIQUITIN.  
DR SMART; SM00213; UBO; 1.  
DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
KW Nuclear protein; Polypeptide.  
FT SITE 48  
FT NECESSARY FOR BRANCHED-CHAIN  
FT MULTIBIQUITIN ADDUCTS.  
FT CONFIGURATION TO ACCEPTOR PROTEINS.  
SQ BINDING 76 76  
SQ SEQUENCE 76 AA; 8507 MW; C53759576F0C8E47 CRC64;  
  
Query Match 30.2%; Score 32; DB 1; Length 76;  
Best Local Similarity 38.1%; Pred. No. 1.6e+02;  
Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Search completed: October 24, 2002, 15:35:32  
Job time : 8.57377 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:24:04 : Search time 24.4426 Seconds  
(without alignments)  
148.629 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFQVEVPGSQHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 107123

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19: \*  
1: sp.archaea: \*  
2: sp.bacteria: \*  
3: sp.fungi: \*  
4: sp.human: \*  
5: sp.invertebrate: \*  
6: sp.mammal: \*  
7: sp.mhc: \*  
8: sp.organelle: \*  
9: sp.phage: \*  
10: sp.plant: \*  
11: sp.virus: \*  
12: sp.virus: \*  
13: sp.vertibrate: \*  
14: sp.unclassified: \*  
15: sp.virus: \*  
16: sp.bacteriophage: \*  
17: sp.archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	41.5	91	10	Q9SYF4
2	41	38.7	91	16	Q9JRI4
3	40	37.7	79	16	Q970B1
4	40	37.7	79	16	Q97P24
5	40	37.7	91	10	Q9M4A4
6	39	36.8	65	10	Q9M4H3
7	39	36.8	77	11	Q9QYX8
8	39	36.8	83	7	Q19274
9	37	34.9	44	10	Q42153
10	37	34.9	45	5	Q36044
11	37	34.9	47	3	Q07188
12	37	34.9	62	5	Q36058
13	37	34.9	62	5	Q36043
14	37	34.9	73	6	Q9GMA0
15	37	34.9	77	2	Q9RJ40
16	37	34.9	79	4	Q96HT2

17	37	34.9	82	3	Q13480	013480 echinodoti
18	36	34.0	48	10	Q9SAT2	Q9sat2 zea mays (m
19	36	34.0	41	2	Q9X9Z0	Q9x9z0 streptomyc
20	36	34.0	50	2	Q48772	Q48772 listeria mo
21	36	34.0	50	2	Q48773	Q48773 listeria mo
22	36	34.0	52	10	Q42009	Q42009 arabidopsis
23	36	34.0	54	10	Q42001	Q42001 arabidopsis
24	36	34.0	72	10	Q9M646	Q9m646 oryza sativ
25	36	34.0	76	10	Q941M2	Q941m2 medicago tr
26	36	34.0	76	10	Q93XW4	Q93xw4 maus domes
27	36	34.0	77	10	Q41570	Q41570 triticum ae
28	36	34.0	77	10	P93379	P93379 nicotiana t
29	36	34.0	78	10	Q24122	Q24122 nicotiana p
30	36	34.0	81	7	Q9MXH8	Q9mxh8 ovbpos mosc
31	36	34.0	92	10	Q42202	Q42202 arabidopsis
32	36	34.0	92	10	Q941N9	Q941n9 alstremeri
33	36	34.0	99	10	Q41431	Q41431 solanum tub
34	36	34.0	100	10	Q9AY08	Q9ay08 eustoma gra
35	35	33.0	27	13	Q90WP2	Q90wp2 fuigu rubrip
36	35	33.0	68	2	Q9F2S1	Q9f2s1 streptomyc
37	35	33.0	73	5	Q25833	Q25833 plasmodium
38	35	33.0	76	3	Q9P3Z3	Q9p3z3 glomus moss
39	35	33.0	77	16	Q9P164	Q9p164 campylobact
40	35	33.0	94	12	Q91F01	Q91f01 cydia pomon
41	35	33.0	99	8	Q37764	Q37764 mytilus edu
42	34.5	32.5	89	2	Q9AKD3	Q9akd3 rickettsia
43	34	32.1	80	5	Q9TW97	Q9tw97 plasmodium
44	34	32.1	80	5	Q9TW76	Q9tw76 plasmodium
45	34	32.1	80	5	Q9U0Q4	Q9u0q4 plasmodium

#### ALIGNMENTS

RESULT 1  
Q9SYF4 PRELIMINARY: PRT: 91 AA.  
AC Q9SYF4;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE F1511.6 PROTEIN.  
GN F1511.6.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,  
RA Araujo R., Chao O., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,  
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006577; AAD25769.1; -  
DR HSSP; P02248; IUBI.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF00240; ubiquitin.2.  
SQ SEQUENCE 91 AA: 10142 MW: E8766823D6450267 CRC64;

Query Match 41.5%; Score 44; DB 10; Length 91;  
Best Local Similarity 47.6%; Pred. No. 6.9;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
QY 1 GATFQVEVPGSQHIDSOKKAI 21  
DB 58 GKTFNLEVKSEETIQOVKNMI 78

```

RESULT 2
Q9JRI4 PRELIMINARY; PRT; 91 AA.
ID Q9JRI4
AC Q9JRI4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN NMA1380.
GN NMA1380 OR NMB1218.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis (serogroup B).
OC Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699, 491;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagers K.M., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Yamashiro Y.,
RA Gall J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AL162755; CAB84626.1; -.
DR TIGR: NMB1218; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 10228 MW; 0C8134C284B66288 CRC64;

Query Match 38.7%; Score 41; DB 16; Length 91;
Best Local Similarity 38.1%; Pred. No. 22;
Matches 8; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATFQVEVPSQHIDSQKAI 21
    ||| ||| : ||| :
Db 61 GATFQVKEVNEQOLDNITRAL 81

RESULT 3
Q97OB1 PRELIMINARY; PRT; 79 AA.
ID Q97OB1
AC Q97OB1;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 1566 FAMILY ELEMENT, ORF1.
GN SP1314.
OS Streptococcus pneumoniae.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AE007441; AAK75537.1; -.
DR TIGR: SP1443; -.
KW Complete proteome.
SQ SEQUENCE 79 AA; 9294 MW; 553CFA03DB89FD6A CRC64;

Query Match 37.7%; Score 40; DB 16; Length 79;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEPGSGHIDSQK 19
    ||| ||| : ||| :
Db 26 VEPGSRFRDSKRR 39

RESULT 4
Q97P24 PRELIMINARY; PRT; 79 AA.
ID Q97P24
AC Q97P24;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 1566 FAMILY ELEMENT, ORF1.
GN SP1443.
OS Streptococcus pneumoniae.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AE007442; AAK75412.1; -.
DR TIGR: SP1314; -.
KW Complete proteome.
SQ SEQUENCE 79 AA; 9260 MW; B5319A03D3235F62 CRC64;

Query Match 37.7%; Score 40; DB 16; Length 79;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEPGSGHIDSQK 19
    ||| ||| : ||| :
Db 26 VEPGSRFRDSKRR 39

RESULT 5
Q9M4A4 PRELIMINARY; PRT; 91 AA.
ID Q9M4A4
AC Q9M4A4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

```

```
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE RESISTANCE GENE HOMOLOGUE (FRAGMENT).
GN MRG-4.
OS Cucumis melo (Muskmelon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=3656;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Garcia-Mas J., van Leeuwen H., Monfort A., de Vicente M.C.,
RA Puigdomenech P., Arus P.;
RT "Cloning and mapping of resistance gene homologues in melon.";
RL Plant Sci. 161:165-172(2001).
DR EMBL: AJ251872; CAB88871.1; -.
DR InterPro: IPR002182; NB-ARC.
DR Pfam: PF00931; NB-ARC; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 91 AA; 10139 MW; FD663CB318C1B881 CRC64;

Query Match
Best Local Similarity 37.7%; Score 40; DB 10; Length 91;
Matches 6; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

OY 4 FQVEVPGSGHIDSOKKAI 21
DB 64 FKMSCPSSDYLDSKRAV 81

RESULT 6
O9M4H3 PRELIMINARY; PRT; 65 AA.
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE METALLOTHIONEIN-LIKE PROTEIN.
GN GRIP24.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC Vitis.
OX NCBI_TaxID=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SHIRAZ; TISSUE=FRUIT;
RC MEDLINE=20177861; PubMed=10712544;
RA Davies C., Robinson S.P.;
RT "Differential screening indicates a dramatic change in mRNA profiles
RT during grape berry ripening. Cloning and characterization of cDNAs
RT encoding putative cell wall and stress response proteins.";
RL Plant Physiol. 122:803-812(2000).
DR EMBL: AJ237990; CAB8630.1; -.
SQ SEQUENCE 65 AA; 6777 MW; B5EA7D8D7B9170D4 CRC64;

Query Match
Best Local Similarity 36.8%; Score 39; DB 10; Length 65;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 AFQVEVPGSGHIDSOK 18
DB 35 ATVMEVPAQHGSGSK 51

RESULT 7
O9QYJ8 PRELIMINARY; PRT; 77 AA.
AC O9QYJ8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
```

```
DE DECAP ACCELERATING FACTOR (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTARST;
RX MEDLINE=20130146; PubMed=10663575;
RA Miwa T., Okada N., Okada H.;
RT "Alternative exon usage in the 3' region of a single gene generates
RT glycosylphosphatidylinositol-anchored and transmembrane forms of rat
RT decay-accelerating factor.";
RL Immunogenetics 51:129-137(2000).
DR EMBL: AB026905; BAA88994.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 77 AA; 8409 MW; B822DC894E684593 CRC64;

Query Match
Best Local Similarity 36.8%; Score 39; DB 11; Length 77;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 5 QVEVPGSGHIDSOKKAI 21
DB 50 EVKVPATQHVPSKTV 66

RESULT 8
O19274 PRELIMINARY; PRT; 83 AA.
AC O19274;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MHC CLASS II ANTIGEN (FRAGMENT).
OS Saguinus oedipus (Cotton-top tamarin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.
OX NCBI_TaxID=9490;
RN [1]
RP SEQUENCE OF 3-30 FROM N.A.
RC STRAIN=B95-8; TISSUE=BLOOD;
RX MEDLINE=97252963; PubMed=9098421;
RA Bidwell J.L., Lu P., Wang Y., Zhou K., Clay T.M., Bontrop R.E.;
RT "DRB, DQA, DOB and DPB nucleotide sequences of Saguinus oedipus B95-
RT 8.";
RL Eur. J. Immunogenet. 21:67-77(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B95-8; TISSUE=BLOOD;
RA de Groot N.N.;
RT Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DL EMBL: AF027966; AAB84042.1; -.
DR InterPro: IPR00353; MHC_II_beta.
DR Pfam: PF00869; MHC_II_beta; 1.
DR ProDom: PD000328; MHC_II_beta; 1.
DR Glycoprotein; MHC II; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 83 AA; 9737 MW; 465CF7063360C9BE CRC64;

Query Match
Best Local Similarity 36.8%; Score 39; DB 7; Length 83;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 2 AFQVEVPGSGHIDSOK 18
DB 40 AVTEIGLPAEHIDSOK 56

RESULT 9
Q42153
```

ID 042153 PRELIMINARY; PRT; 44 AA.  
 AC 042153;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RELATIVE TO UBIQUITIN, PARATHYROID.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;  
 RA Lebas M., Regad F., Lescure B.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z26872; CAA81485.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 SQ SEQUENCE 44 AA; 4860 MW; DB9E08BD8C673EE9 CRC64;

Query Match 34.9%; Score 37; DB 10; Length 44;  
 Best Local Similarity 33.3%; Pred. No. 47;  
 Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKAI 21  
 ID 036044 PRELIMINARY; PRT; 45 AA.  
 AC 036044;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE UBIQUITIN (FRAGMENT).  
 OS Mastigamoeba balamuthi.  
 OC Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 ON NCBI\_TaxID=108607;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Woestmann C., Hackstein J., Bakker-Grünwald T.;  
 RT "Ubiquitin sequences and polyubiquitin genes in phreatamoeba balamuthi  
 and Psalteriomonas lanterna: Phylogenetic implications."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ000658; CAA04206.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR00348; UBIQUITIN.  
 DR SMART; SM00213; UBO; 1.  
 DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 45 45  
 SQ SEQUENCE 45 AA; 5090 MW; 370BA35AD569F4A4 CRC64;

Query Match 34.9%; Score 37; DB 5; Length 45;  
 Best Local Similarity 42.9%; Pred. No. 48;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKAI 21  
 ID 007188 PRELIMINARY; PRT; 47 AA.  
 AC 007188;  
 DT 01-JAN-1998 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 09, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RELATIVE TO UBIQUITIN, PARATHYROID.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;  
 RA Lebas M., Regad F., Lescure B.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z26872; CAA81485.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 SQ SEQUENCE 44 AA; 4860 MW; DB9E08BD8C673EE9 CRC64;

RESULT 11  
 ID 007188 PRELIMINARY; PRT; 47 AA.  
 AC 007188;  
 DT 01-JAN-1998 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 09, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RELATIVE TO UBIQUITIN, PARATHYROID.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;  
 RA Lebas M., Regad F., Lescure B.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z26872; CAA81485.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 SQ SEQUENCE 44 AA; 4860 MW; DB9E08BD8C673EE9 CRC64;

DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE UBIQUITIN.  
 GN UBI4 AND YIL039C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 ON NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85061630; PubMed=6095120;  
 RA Osekaynak E., Finley D., Varshavsky A.;  
 RT "The yeast ubiquitin gene: head-to-tail repeats encoding a  
 RT polyubiquitin precursor protein."  
 RL Nature 312:663-666(1984).  
 DR EMBL; X01473; CAA25705.1; -  
 DR HSSP; P02248; IUBI.  
 DR SGD; S0003962; UBI4.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00213; UBO; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 47 AA; 5220 MW; BD2AAE57CE57C8E9 CRC64;

Query Match 34.9%; Score 37; DB 3; Length 47;  
 Best Local Similarity 42.9%; Pred. No. 51;  
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKAI 21  
 ID 036058 PRELIMINARY; PRT; 62 AA.  
 AC 036058;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE UBIQUITIN (FRAGMENT).  
 OS Mastigamoeba balamuthi.  
 OC Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 ON NCBI\_TaxID=108607;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Woestmann C., Hackstein J., Bakker-Grünwald T.;  
 RT "Ubiquitin sequences and polyubiquitin genes in phreatamoeba balamuthi  
 and Psalteriomonas lanterna: Phylogenetic implications."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ000656; CAA04204.1; -  
 DR EMBL; AJ000655; CAA04203.1; -  
 DR EMBL; AJ000659; CAA04207.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR00348; UBIQUITIN.  
 DR SMART; SM00213; UBO; 1.  
 DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 KW Nuclear protein.  
 FT NON\_TER 1 1  
 FT NON\_TER 62 62  
 SQ SEQUENCE 62 AA; 6978 MW; 7C14ED7613B19D20 CRC64;

Query Match 34.9%; Score 37; DB 5; Length 62;  
 Best Local Similarity 42.9%; Pred. No. 70;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKAI 21  
 ID 007188 PRELIMINARY; PRT; 47 AA.  
 AC 007188;  
 DT 01-JAN-1998 (TREMBlrel. 01, Created)  
 DT 01-JAN-1998 (TREMBlrel. 09, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RELATIVE TO UBIQUITIN, PARATHYROID.  
 OS Arabidopsis thaliana (mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CELL SUSPENSION CULTURE OF ECOTYPE COLUMBIA;  
 RA Lebas M., Regad F., Lescure B.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z26872; CAA81485.1; -  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 SQ SEQUENCE 44 AA; 4860 MW; DB9E08BD8C673EE9 CRC64;

```

Db          27  QVENPASHVDNE 39

RESULT 15
O9RJ40
ID 09RJ40 PRELIMINARY; PRT; 77 AA.
AC 09RJ40:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 8.4 KDA PROTEIN.
SC 18.13.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC Oliver K., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RX Redenbach M., Kleser H.M., Denapite D., Eichner A., Cullum J.,
RX Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL132644; CAB59444.1; -.
KW Hypothetical protein.
SQ SEQUENCE 77 AA; 8352 MW; 86EA16403C7B7A1 CRC64;

Query Match 34.9%; Score 37; DB 2; Length 77;
Best Local Similarity 42.9%; Fred. No. 89;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHI 14
| ||| | | : | :
Db 12 GATFEVHAQGAHV 25

Search compile d: October 24, 2002, 15:36:52
Job time : 25.4426 secs

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:17 ; Search time 33.7377 Seconds  
(without alignments)  
69.138 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATPQVEVPSQHIDSKKAI 21

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	21	AAV87462	Cholera toxin B/en
2	106	100.0	103	AAW04857	Synthetic cholera
3	106	100.0	103	AAW04939	Heat labile entero
4	106	100.0	103	AAW06606	Cholera toxin B su
5	106	100.0	103	AAW06607	Cholera toxin B su
6	106	100.0	103	AAW08088	Amino acid sequenc
7	106	100.0	103	AAW62367	V. cholera cholera
8	106	100.0	103	AAW62370	V. cholera cholera
9	106	100.0	105	AAW62369	Cholera toxin B-su
10	106	100.0	118	AAW04163	E. coli LTB protei
11	106	100.0	123	AAW62374	

12	106	100.0	124	10	AAW93561	B subunit of the h
13	106	100.0	124	17	AAW06605	Cholera toxin B su
14	106	100.0	124	19	AAW59770	Amino acid sequenc
15	106	100.0	124	21	AAV96652	Plant-optimized E.
16	106	100.0	124	22	AAW65992	Cholera toxin B su
17	106	100.0	124	22	AAW62359	V. cholera strain
18	106	100.0	124	22	AAW62363	V. cholera cholera
19	106	100.0	124	22	AAW62368	V. cholera cholera
20	106	100.0	124	22	AAW62376	E. coli LTB protei
21	106	100.0	124	22	AAW62378	E. coli LTB protei
22	106	100.0	125	22	AAW62372	E. coli LTB protei
23	106	100.0	131	11	AAW04825	LTB-CTB fusion pro
24	106	100.0	138	15	AAW50227	Sequence of LT-B-M
25	106	100.0	170	20	AAW94082	LTB-CTP fusion pro
26	106	100.0	380	22	AAW00507	E. coli heat-labli
27	106	100.0	382	22	AAW00506	E. coli heat-labli
28	106	100.0	405	12	AAW11272	HSV-1 antigen/heat
29	103	97.2	103	6	AAW50340	Sequence of sub-un
30	101	95.3	21	21	AAW87463	E. coli heat labil
31	101	95.3	93	16	AAW72545	ADP-ribosylating t
32	101	95.3	93	20	AAW41816	Escherichia coli v
33	101	95.3	93	20	AAW95226	E. coli heat-labli
34	101	95.3	93	21	AAW68365	Heat labile toxin
35	101	95.3	93	22	AAW66239	E. coli verotoxin-1
36	101	95.3	103	22	AAW62379	E. coli LTB protei
37	101	95.3	123	22	AAW62377	E. coli LTB protei
38	101	95.3	124	22	AAW62373	E. coli LTB protei
39	101	95.3	124	22	AAW62375	E. coli LTB protei
40	101	95.3	124	22	AAW62380	E. coli LTB protei
41	101	95.3	134	22	AAW73241	Recombinant exotox
42	101	95.3	142	22	AAW73242	Recombinant exotox
43	101	95.3	155	22	AAW73243	Recombinant exotox
44	101	95.3	163	22	AAW73244	Recombinant exotox
45	101	95.3	371	20	AAW01300	Labile toxin (LT-B

## ALIGNMENTS

RESULT 1	AAW87462	standard; peptide; 21 AA.
ID	AAW87462:	
AC	AAW87462:	
XX		
DT	03-JUL-2000	(first entry)
XX		
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.	
XX		
KW	Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;	
KW	beta-4-alpha-2 loop; GM-1 ganglioside receptor; Immunomodulation;	
KW	adjuvant; immune disorder; diarrhoea.	
XX		
OS	Vibrio cholerae.	
XX	Escherichia coli.	
PN	WO200014114-A1.	
XX		
PD	16-MAR-2000.	
XX		
PF	07-SEP-1999;	99WO-GB02970.
XX		
PR	07-SEP-1998;	98GB-0019484.
XX		
PA	(UYBR-) UNIV BRISTOL.	
XX		
PI	Williams NA, Hirst TR;	
XX		
DR	WPI: 2000-256943/22.	
XX		
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as	
PT	immunomodulators and for treating diarrhoea and which do not bind the	
PT	glycolipid receptor GM-1 -	

XX XX Disclosure; Page 15; 62pp; English.  
 PS XX  
 CC The invention relates to peptide fragments of the *Escherichia coli* heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from *Vibrio cholerae* which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (CtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.  
 CC XX  
 SQ Sequence 21 AA;  
 Query Match 100.0%; Score 106; DB 21; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATFOVEVPGSQHIDSOKKAI 21  
 Db 1 GATFOVEVPGSQHIDSOKKAI 21  
 RESULT 2  
 AAM04857  
 ID AAM04857 standard; Protein; 103 AA.  
 AC AAM04857;  
 XX  
 DT 21-FEB-1997 (first entry)  
 XX  
 DE Synthetic cholera toxin B subunit.  
 XX  
 KW Bordetella pertussis; whooping cough; recombinant construct;  
 KW cholera toxin B subunit; enzyme; antigen; immunogen; allergen;  
 KW enzyme inhibitor; hormone; lymphokine; immunoglobulin; toxin;  
 KW structural protein; receptor; heterologous gene; leader; promoter.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9626282-A1.  
 XX  
 PD 29-AUG-1996.  
 XX  
 PF 23-FEB-1996; 96WO-CA00107.  
 XX  
 PR 23-FEB-1995; 95US-0393334.  
 XX  
 PA (CONN-) CONNANUGHT LAB LTD.  
 XX  
 PI Klein MH, Loosmore SM, Yacoub RK, Zealey GR;  
 XX  
 DR WPI: 1996-425088/42.  
 DR N-PSDB; AAT38038.  
 XX  
 XX Recombinant constructs for expressing and opt. secreting proteins in  
 PT Bordetella - comprise Bordetella promoter coupled to non-Bordetella,  
 PT esp. cholera B toxin, gene or coupled to non-Bordetella leader and  
 PT gene of interest  
 XX

PS Example 1; Figure 1; 61pp; English.  
 XX  
 CC Recombinant constructs comprising a promoter functional in  
 CC Bordetella operatively linked to a heterologous gene or a non-  
 CC Bordetella leader sequence for secretion of a gene product which may  
 CC or may not be of Bordetella origin, can be used for the expression  
 CC in Bordetella of enzymes, antigens, immunogens, allergens, enzyme  
 CC inhibitors, hormones, lymphokines, immunoglobulins or their  
 CC fragments, toxins, mammalian proteins, structural proteins or  
 CC receptors. The Bordetella strains are particularly engineered to  
 CC express the cholera toxin B subunit (this sequence). The promoters  
 CC used in the constructs are selected from the Bordetella pertussis  
 CC tox, fha promoters or the high molecular weight (hmw) outer membrane  
 CC promoter of non typable *Haemophilus influenzae*; leaders used in the  
 CC constructs are selected from the cholera toxin B leader (CTB-L), the  
 CC pertussis toxin subunit S1 leader (ST-L) and the pertussin pertactin  
 CC leader (PTN-L); and genes used in the constructs are selected from a  
 CC novel synthetic cholera toxin B gene (ctb) and the hmw1 and hmw2  
 CC genes of *Haemophilus influenzae*.  
 CC XX  
 SQ Sequence 103 AA;  
 Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATFOVEVPGSQHIDSOKKAI 21  
 Db 45 GATFOVEVPGSQHIDSOKKAI 65  
 RESULT 3  
 AAR94939  
 ID AAR94939 standard; Protein; 103 AA.  
 AC AAR94939;  
 XX  
 DT 31-OCT-1996 (first entry)  
 XX  
 DE Heat labile enterotoxin B subunit (LT-B) *E. coli*.  
 XX  
 KW Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;  
 KW adjuvant; immunisation.  
 XX  
 OS *Escherichia coli*.  
 XX  
 PN WO9612801-A1.  
 XX  
 PD 02-MAY-1996.  
 XX  
 PF 24-OCT-1995; 95WO-US13376.  
 XX  
 PR 24-OCT-1994; 94US-0328716.  
 XX  
 PA (TULA ) TULANE EDUCATIONAL FUND.  
 PA (TEXA ) UNIV TEXAS A & M SYSTEM.  
 XX  
 PI Arntzen CJ, Clements JD, Haq TA, Mason HS;  
 XX  
 DR WPI: 1996-230602/23.  
 DR N-PSDB; AAT18799, AAT18800.  
 XX  
 XX Transgenic plants contg. *E. coli* heat labile enterotoxin subunits -  
 PT used as oral vaccines for animals which consume the plant  
 PT  
 PS Disclosure; Page 100-101; 130pp; English.  
 XX  
 CC A transgenic plant comprising or expressing a DNA sequence encoding  
 CC an immunogenic agent can be used as an oral vaccine for animals.  
 CC The vaccine is administered by the oral consumption of the plant and  
 CC provides the first known functional method for immunising animals  
 CC using transgenic plants, where the plants express bacterial antigens  
 CC that act as both immunogens and adjuvants. The method provides an



CC inexpensive production and delivery system for such antigens to  
 CC animals. This is the LT-B Escherichia coli toxin subunit and its  
 CC coding sequence was used in the construction of such a transgenic  
 CC plant. The immunogenic agent preferably comprises the LT-B or CT-B  
 CC (cholera toxin B subunit) or optionally LT-A or CT-A.  
 CC

SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKAI 21  
 |||||  
 Db 45 GATFOVEVPGSQHIDSQKAI 65

#### RESULT 4

AAW06606  
 ID AAW06606 standard; Protein; 103 AA.

AC AAW06606;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

OS Chimeric - Vibrio cholerae.  
 OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers

FT MISC-difference 1 /label= substitution

FT /note= "wild-type Thr replaced by Ala"

FT MISC-difference 94 /label= substitution

FT /note= "wild-type His replaced by Asn"

FT MISC-difference 95 /label= substitution

FT /note= "wild-type Ala replaced by Ser"

XX WO9634893-A1.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI: 1996-506108/50.

XX N-PSDB; AAT43576.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit

XX hybrid protein - opt. fused to immunogenic sequence for use in

XX vaccines against enterotoxin-induced illness

XX Claim 3; Page -: 32pp; English.

XX AAW06606 is a mature cholera toxin B subunit (CTB)/heat labile

XX enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,

XX certain amino acids (aa) were replaced with corresponding aa from

XX heat-labile enterotoxin B subunit (LTB). The specific amino acid

XX substitutions impart LTB-specific epitope characteristics to

CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAW06605).

SQ Sequence 103 AA;  
 Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKAI 21  
 |||||  
 Db 45 GATFOVEVPGSQHIDSQKAI 65

#### RESULT 5

AAW06607  
 ID AAW06607 standard; Protein; 103 AA.

AC AAW06607;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit/heat labile enterotoxin B subunit hybrid.

KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;  
 KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;  
 KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;  
 KW vomiting; food poisoning.

OS Chimeric - Vibrio cholerae.  
 OS Chimeric - Enterotoxigenic Escherichia Coli.

XX Key Location/Qualifiers

FT MISC-difference 1..25 /label= substitution

FT /note= "the first 25 amino acids of mature

FT wild-type cholera toxin B subunit are

FT replaced with the first 25 amino acids

FT of mature enterotoxin B subunit"

XX WO9634893-A1.

XX 07-NOV-1996.

XX 02-MAY-1996; 96WO-SE00570.

XX 05-MAY-1995; 95SE-0001682.

XX (HOLM/) HOLMGREN J.

XX (LEBE/) LEBENS M R.

XX Holmgren J, Lebens MR;

XX WPI: 1996-506108/50.

XX N-PSDB; AAT43577.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit

XX hybrid protein - opt. fused to immunogenic sequence for use in

XX vaccines against enterotoxin-induced illness

XX Claim 4; Page -: 32pp; English.

XX AAW06607 is a mature cholera toxin B subunit (CTB)/heat labile

XX enterotoxin B subunit (LTB) hybrid molecule. In mature CTB protein,

XX certain amino acids (aa) were replaced with corresponding aa from

XX heat-labile enterotoxin B subunit (LTB). The specific amino acid

CC substitutions impart LTB-specific epitope characteristics to  
 CC immunogenic mature CTB. The hybrid molecules have increased  
 CC cross-reactivity and are suitable for a broad spectrum vaccine to  
 CC protect against enterotoxigenic illness. Immunogenic proteins  
 CC comprising the hybrid molecules can be used to treat, or in a  
 CC vaccine to prevent, enterotoxin-induced illness, e.g. diarrhoea  
 CC and vomiting, in humans and animals.  
 CC Note - This sequence does not appear in the specification, it is  
 CC a claimed mutant sequence of mature cholera toxin B subunit (see  
 CC AAM06605).

XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 17; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21  
 DB 45 GATFOVEVPGSQHIDSOKKAI 65  
 |||||||||||||||||||

RESULT 6  
 ID AAM80808 standard; protein: 103 AA.  
 AC AAM80808;  
 DT 29-JAN-1999 (first entry)  
 XX  
 XX Amino acid sequence of the wild type choreia toxin B subunit.  
 DE  
 XX Cholera toxin B subunit; nontoxic subunit; adjuvant; coadministration;  
 KM antigen; bird; animal; mucosal; vaccine.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN W09845324-A1.  
 PD 15-OCT-1998.  
 XX  
 PF 03-APR-1998; 98WO-US06725.  
 XX  
 PR 04-APR-1997; 97US-0043410.  
 XX  
 XX (KIYO/) KIYONO H.  
 PA (MCGH/) MCGHEE J R.  
 PA (TAKE/) TAKEDA Y.  
 PA (UABR-) UAB RES FOUND.  
 PA (YAMA/) YAMAMOTO S.  
 XX  
 PI Kiyono H, Mcghee JR, Takeda Y, Yamamoto S;  
 DR WPI; 1998-594478/50.  
 XX  
 PT New mutant cholera toxin selected from a group comprising nontoxic  
 PT subunits/derivatives - effective as an adjuvant when coadministered  
 PT with an antigen to birds and mammals  
 XX  
 PS Disclosure; Fig 1B; 43pp; English.

XX This is the amino acid sequence of the cholera toxin B subunit used in  
 CC the method of the invention involving the use of nontoxic subunits as  
 CC an effective adjuvant in coadministration of an antigen to birds  
 CC and animals. In addition to the use of the toxin as an mucosal  
 CC adjuvant, it also provides a vaccine comprising the toxin, an  
 CC immunogenic amount of an antigen, and a pharmaceutically acceptable  
 CC carrier. The toxin can be used with single/multiple vaccines, and it  
 CC enables the possibility for commercial mucosal adjuvants for use in  
 CC humans, since these are more effective and safer than vaccines  
 CC administered subcutaneously.

XX Sequence 103 AA;

Query Match 100.0%; Score 106; DB 19; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21  
 DB 45 GATFOVEVPGSQHIDSOKKAI 65  
 |||||||||||||||||||

RESULT 7  
 ID AAB62367 standard; Protein: 103 AA.  
 AC AAB62367;  
 DT 29-JUN-2001 (first entry)  
 XX  
 XX V. cholera cholera toxin B (CTB) protein variant (GI 1421511).  
 DE  
 XX AAB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 KM immunogen; autoimmune disease; neuron demyelinating disease; rCTB;  
 KM infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;  
 KM antidiabetic; neuroprotective; vaccine; variant.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN W0200127144-A2.  
 PD 19-APR-2001.  
 XX  
 PF 05-OCT-2000; 2000WO-US27607.  
 XX  
 PR 08-OCT-1999; 99US-0158561.  
 XX  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 PA  
 XX Handley HH, Haaparanta T, Ewalt KL;  
 PI WPI; 2001-281974/29.  
 DR  
 XX  
 PT Recombinant AAB5B subunit protein comprising a mutation that alters the  
 PT number residues available for chemical modification, useful for  
 PT covalently linking to an immunogen or vaccine which can be used for  
 PT treating autoimmune diseases -  
 XX  
 PS Disclosure; Page 67-68; 78pp; English.

XX The invention relates to a recombinant AAB5B subunit protein comprising at  
 CC least one mutation, where the mutation alters the number of amino acid  
 CC residues available for chemical modification as compared to a wild type  
 CC AAB5B subunit protein, and where the recombinant protein retains an  
 CC effective target ligand binding affinity. A recombinant AAB5B subunit  
 CC protein such as cholera toxin B protein (CTB) can be specifically  
 CC covalently linked at lysines or cysteines to an immunogen or vaccine.  
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron  
 CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the  
 CC parasite infection. The present sequence represents a CTB variant  
 CC protein, NCBI GenBank No. GI 1421511.

XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 106; DB 22; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21  
 DB 45 GATFOVEVPGSQHIDSOKKAI 65  
 |||||||||||||||||||

```

RESULT 8
AAB62370
ID AAB62370 standard; Protein; 103 AA.
XX
AC AAB62370;
XX
DT 29-JUN-2001 (first entry)
XX
DE V. cholera cholera toxin B (CTB) protein variant (GI 1421525).
XX
KW AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB;
KW infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
KW antidiabetic; neuroprotective; vaccine; variant.
XX
OS Vibrio cholerae.
XX
PN WO200127144-A2.
XX
PD 19-APR-2001.
XX
PF 05-OCT-2000; 2000WO-US27607.
XX
PR 08-OCT-1999; 99US-0158561.
XX
PA (ACT1-) ACTIVE BIOTECH AB.
XX
PI Handley HH, Haaparanta T, Ewalt KL;
XX
DR WPI; 2001-281974/29.
XX
PT Recombinant AB5B subunit protein comprising a mutation that alters the
PT number residues available for chemical modification, useful for
PT covalently linking to an immunogen or vaccine which can be used for
PT treating autoimmune diseases -
XX
PS Disclosure; Page 69; 78pp; English.
XX
CC The invention relates to a recombinant AB5B subunit protein comprising at
CC least one mutation, where the mutation alters the number of amino acid
CC residues available for chemical modification as compared to a wild type
CC AB5B subunit protein, and where the recombinant protein retains an
CC effective target ligand binding affinity. A recombinant AB5B subunit
CC protein such as cholera toxin B protein (CTB) can be specifically
CC covalently linked at lysines or cysteines to an immunogen or vaccine.
CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g.
CC parasitic infection. The present sequence represents a CTB variant
CC protein, NCBI Genbank No. GI 1421525 (Ogawa 41 K35D).
XX
SQ Sequence 103 AA;
XX
Query Match 100.0%; Score 106; DB 22; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATFOVEVPGSOHIDSOKKAI 21
| | | | | | | | | | | | | | | | | | | | | |
Db 45 GATFOVEVPGSOHIDSOKKAI 65

```

```

KW AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;
KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB;
KW infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;
KW antidiabetic; neuroprotective; vaccine; variant.
XX
OS Vibrio cholerae.
XX
PN WO200127144-A2.
XX
PD 19-APR-2001.
XX
PF 05-OCT-2000; 2000WO-US27607.
XX
PR 08-OCT-1999; 99US-0158561.
XX
PA (ACT1-) ACTIVE BIOTECH AB.
XX
PI Handley HH, Haaparanta T, Ewalt KL;
XX
DR WPI; 2001-281974/29.
XX
PT Recombinant AB5B subunit protein comprising a mutation that alters the
PT number residues available for chemical modification, useful for
PT covalently linking to an immunogen or vaccine which can be used for
PT treating autoimmune diseases -
XX
PS Disclosure; Page 68-69; 78pp; English.
XX
CC The invention relates to a recombinant AB5B subunit protein comprising at
CC least one mutation, where the mutation alters the number of amino acid
CC residues available for chemical modification as compared to a wild type
CC AB5B subunit protein, and where the recombinant protein retains an
CC effective target ligand binding affinity. A recombinant AB5B subunit
CC protein such as cholera toxin B protein (CTB) can be specifically
CC covalently linked at lysines or cysteines to an immunogen or vaccine.
CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune
CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron
CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g.
CC parasitic infection. The present sequence represents a CTB variant
CC protein, NCBI Genbank No. GI 2781121 (Ogawa 41).
XX
SQ Sequence 105 AA;
XX
Query Match 100.0%; Score 106; DB 22; Length 105;
Best Local Similarity 100.0%; Pred. No. 3.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATFOVEVPGSOHIDSOKKAI 21
| | | | | | | | | | | | | | | | | | | | | |
Db 46 GATFOVEVPGSOHIDSOKKAI 66

```

```

RESULT 10
AAR04163
ID AAR04163 standard; protein; 118 AA.
XX
AC AAR04163;
XX
DT 10-SEP-1990 (first entry)
XX
DE Cholera Toxin B-subunit.
XX
KW cholera toxin B-subunit; hybrid protein; heterologous IGA active antigen.
XX
OS synthetic.
XX
FH Key Location/Qualifiers
FT msc_difference 18..18
FT /*label= His or Tyr
FT 1..11
FT region
FT /*note= absent from mature protein
XX

```

XX W09003437-A.  
 XX 05-APR-1990.  
 XX 27-SEP-1989; 89WO-0000495.  
 XX 27-SEP-1988; 88FR-0012627.  
 XX (UYLI-) L/UNIVERSITE DE L'ETAT A LIEGE.  
 XX L/Hoir C, Renard A, Martial J;  
 XX WPI: 1990-132273/17.  
 XX N-NSDB; 004046.  
 XX New hybrid protein, useful in vaccines -  
 XX contains cholera toxin b subunit and heterologous IgA active  
 XX antigenic sequence.  
 XX Disclosure; : pp; French.  
 XX Mature cholera toxin B-subunit is obtained when the signal peptide is  
 XX cleaved off. There is an ochre codon at position 343-5; the sequence  
 XX downstream from it is part of a plasmid.  
 XX Sequence 118 AA;  
 SO  
 Query Match 100.0%; Score 106; DB 11; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GATFOVEVPGSQHIDSQKKAI 21  
 ||||||||||||||||||||  
 DB 56 GATFOVEVPGSQHIDSQKKAI 76  
 RESULT 11  
 AAB62374  
 ID AAB62374 standard; Protein: 123 AA.  
 XX AAB62374;  
 XX 29-JUN-2001 (first entry)  
 XX E. coli LTB protein variant (GI: 1395122).  
 XX AAB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 XX immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;  
 XX infection; parasitic; immunosuppressive; antiarthritic; anti-rheumatic;  
 XX antididactic; neuroprotective; vaccine; heat labile toxin B; variant.  
 XX Escherichia coli.  
 XX W0200127144-A2.  
 XX 19-APR-2001.  
 XX 05-OCT-2000; 2000WO-US27607.  
 XX 08-OCT-1999; 99US-0158561.  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 XX Handley HH, Haaparanta T, Ewalt KL;  
 XX WPI: 2001-281974/29.  
 XX Recombinant ABB5 subunit protein comprising a mutation that alters the  
 XX number residues available for chemical modification, useful for  
 XX covalently linking to an immunogen or vaccine which can be used for  
 XX treating autoimmune diseases -

PS Disclosure; Page 70; 78pp; English.  
 XX The invention relates to a recombinant ABB5 subunit protein comprising at  
 XX least one mutation, where the mutation alters the number of amino acid  
 XX residues available for chemical modification as compared to a wild type  
 XX ABB5 subunit protein, and where the recombinant protein retains an  
 XX effective target ligand binding affinity. A recombinant ABB5 subunit  
 XX protein such as cholera toxin B protein (CTB) can be specifically  
 XX covalently linked at lysines or cysteines to an immunogen or vaccine.  
 XX Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune  
 XX diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron  
 XX demyelinating diseases) and diabetes. The rCTB or other B subunits of the  
 XX invention can also be used to induce tolerance to infection, e.g.  
 XX parasitic infection. The present sequence represents a variant of the  
 XX E. coli heat labile toxin B (LTB) protein.  
 XX Sequence 123 AA;  
 SO  
 Query Match 100.0%; Score 106; DB 22; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GATFOVEVPGSQHIDSQKKAI 21  
 ||||||||||||||||||||  
 DB 66 GATFOVEVPGSQHIDSQKKAI 86  
 RESULT 12  
 AAP93561  
 ID AAP93561 standard; protein: 124 AA.  
 XX AAP93561;  
 XX 06-JUN-1990 (first entry)  
 XX B subunit of the heat-labile enterotoxin (LT-B) derived from E. coli.  
 XX B subunit; heat-labile enterotoxin; LT-B; Escherichia coli; malaria;  
 XX cilioporozyte protein; fusion protein; live recombinant vaccine;  
 XX Salmonella; epitope.  
 XX Escherichia coli.  
 XX Key Location/Qualifiers  
 XX Peptide 1..22  
 XX /note="Signal peptide"  
 XX Protein 23..124  
 XX /note="Mature LT-B"  
 XX W08902924-A.  
 XX 06-APR-1989.  
 XX 30-SEP-1988; 88WO-US03376.  
 XX 02-OCT-1987; 87US-0104735.  
 XX (PRAX-) PRAXIS BIOLOGICS IN.  
 XX Brey RN, Majarian WR, Pillai S, Hockmeyer WT;  
 XX WPI: 1989-114399/15.  
 XX N-PSDB; AAN90747.  
 XX Live recombinant vaccine for malaria -  
 XX comprising attenuated entero-invasive bacterium contg. DNA  
 XX encoding epitope of malaria parasite  
 XX Fig 3; p. 3/17; 105pp; English.  
 XX In the patent, the DNA encoding LT-B is expressed as part of a fusion  
 XX protein with an epitope of a malaria parasite, eg Region I or Region II  
 XX or a repeat region of circumsporozoite protein antigen (CS) (AAP93560)

CC from *Plasmodium berghiei*. Pref. the fusion gene is inserted into  
 CC attenuated *Salmonella enteritidis* under the left promoter control of  
 CC *lambda*. Such bacteria can multiply in the host without causing disease or  
 CC disorder and express CS that will induce a protective immune response  
 CC against malaria and can be used in vaccines. Such vaccines can be  
 CC multivalent.

XX Sequence 124 AA;

Query Match 100.0%; Score 106; DB 10; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKRAI 21  
 |||||  
 DB 66 GATFOVEVPGSQHIDSQKRAI 86

RESULT 13  
 AAM06605

ID AAM06605 standard; Protein; 124 AA.

AC AAM06605;

DT 06-AUG-1997 (first entry)

DE Cholera toxin B subunit, used for hybrid immunogenic toxin production.

KW Heat labile; tolerant; stable; hybrid; cholera toxin B subunit; CTB;

KW heat labile enterotoxin B subunit; LTB; vaccine; immunisation;

KW immunogen; antigen; broad spectrum; cross reactive; diarrhoea;

OS vomiting; food poisoning.

XX *Vibrio cholerae*.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= sig\_peptide

PN WO9634893-A1.

PD 07-NOV-1996.

PF 02-MAY-1996; 96WO-SE00570.

PR 05-MAY-1995; 95SE-0001682.

PA (HOLM/) HOLMGREN J.

PI (LEBE/) LEBENS M R.

DR Holmgren J, Lebens MR;

DR MPI; 1996-506108/50.

DR N-PSDB; AAT43575.

XX Heat-labile enterotoxin B sub-unit and cholera toxin B sub-unit

PT hybrid protein - opt. fused to immunogenic sequence for use in

PT vaccines against enterotoxin-induced illness

PS Disclosure; Fig 1; 32pp; English.

XX AAM06605 is the full (including the signal peptide) length cholera  
 CC toxin B subunit (CTB), this sequence is described as unpublished in  
 CC the specification. The mature CTB protein was used to create hybrid  
 CC mutants, in which certain amino acids (aa) of CTB were replaced with  
 CC corresponding aa from heat-labile enterotoxin B subunit (LTB), see  
 CC AAM06606 and AAM06607. The specific amino acid substitutions impart  
 CC LTB-specific epitope characteristics to immunogenic mature CTB. The  
 CC hybrid molecules have increased cross-reactivity and are suitable  
 CC for a broad spectrum vaccine to protect against enterotoxigenic  
 CC illness. Immunogenic proteins comprising the hybrid molecules can be

CC used to treat, or in a vaccine to prevent, enterotoxin-induced illness,  
 CC e.g. diarrhoea and vomiting, in humans and animals.

XX Sequence 124 AA;

Query Match 100.0%; Score 106; DB 17; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKRAI 21  
 |||||  
 DB 66 GATFOVEVPGSQHIDSQKRAI 86

RESULT 14  
 AAM59770

ID AAM59770 standard; Protein; 124 AA.

AC AAM59770;

DT 12-OCT-1998 (first entry)

DE Amino acid sequence of *E. coli* LTB.

KW Beta-subunit of heat labile enterotoxin; LTB; fusion protein; vaccine;

KW immunogen; antigen; inhibitor; fertility; follicle stimulating hormone;

KW FSH; sperm; ova; immune response.

XX *Escherichia coli*.

PN WO9821344-A1.

PD 22-MAY-1998.

PF 12-NOV-1997; 97WO-US20584.

PR 12-NOV-1996; 96US-0747410.

PA (UNMS ) UNIV MICHIGAN STATE.

PI Bagdasarian M, Ireland J;

DR MPI; 1998-297947/26.

DR N-PSDB; AAV41573.

XX New nucleic acid encoding fusion of antigenic peptide and

PT enterotoxin sub-unit - useful as vaccinating immunogen, particularly

PT for increasing animal fertility by inducing antibodies against

PT inhibin

XX Disclosure; Fig 9; 56pp; English.

XX This is the amino acid sequence of *Escherichia coli* beta-subunit of  
 CC heat labile enterotoxin (LTB). It is used in the method of the  
 CC invention to create fusion proteins which are useful as vaccinating  
 CC immunogens. The fusion proteins are useful in vaccines, specifically  
 CC where the antigenic peptide is an inhibin fragment for increasing the  
 CC fertility of an animal (by increasing levels of follicle stimulating  
 CC hormone (FSH) or production of sperm or ova), but more generally for  
 CC inducing an immune response against the antigenic peptide. Vaccines  
 CC are particularly administered orally, e.g. fusion protein is expressed  
 CC in edible plants or animals.

XX Sequence 124 AA;

Query Match 100.0%; Score 106; DB 19; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKRAI 21  
 |||||  
 DB 66 GATFOVEVPGSQHIDSQKRAI 86

RESULT 15  
 ID AAY96652 standard; Protein; 124 AA.  
 AC AAY96652;

DT 26-SEP-2000 (first entry)

DE Plant-optimized E. coli heat labile toxin B subunit.

KM Heat-labile toxin; LT-A; LT-B; mutant; transgenic plant; vaccine; oral;  
 KW adjuvant; anti-bacterial.

OS Escherichia coli.  
 OS Synthetic.

Key	Location/Qualifiers
FT Peptide	1..21
FT	/label= signal_peptide
FT Protein	22..124
FT	/label= mature_protein

PN W0200037609-A2.

PD 29-JUN-2000.

PF 22-DEC-1999; 99MO-US30747.

PR 22-DEC-1998; 98US-0113507.

PA (BOYC-) BOYCE THOMPSON INST PLANT RES.  
 PA (MASO/) MASON H S.

PA (ARNT/) ARNTZEN C J.

PI Mason HS, Arntzen CJ;

DR WPI; 2000-442653/38.

DR N-PSDB; AAA51156.

PT New polynucleotides encoding LT-A or CT-A polypeptides for the  
 transformation of plant cells, useful in immunogenic compositions to  
 elicit immune responses in animals

PS Claim 4; Fig 5A-B; 103pp; English.

CC This synthetic Escherichia coli heat-labile toxin (LT) B subunit (LT-B)  
 is encoded by a plant-codon optimized CDNA. The CDNA sequence contains  
 plant-preferred codons and eliminates sequence motifs associated with  
 spurious mRNA processing. The second codon is changed from AAT encoding  
 Asn to GTG encoding Val, in order to create a NcoI restriction site at  
 the 5' end. Novel polynucleotides encode a mutant LT-A polypeptide or a  
 mutant Vibrio cholerae cholera toxin (CT) A subunit (CT-A) polypeptide,  
 which have reduced enzyme activity as compared to the wild-type LT-A or  
 CT-A polypeptide and where at least one of the codons is altered to a  
 plant preferred codon. The polynucleotide further comprises a nucleic  
 acid sequence encoding LT B subunit (LT-B) or a CT B subunit (CT-B). The  
 polynucleotides are useful for the transformation of plant cells for the  
 production of transgenic plants to produce edible vaccines, especially  
 oral vaccines in transgenic plants for the prophylactic or therapeutic  
 treatment against E. coli or V. cholerae. The mutant polypeptides are  
 also useful as adjuvants.

SQ Sequence 124 AA;

Query Match 100.0%; Score 106; DB 21; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 4e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHTDSOKKAI 21  
 ||||||||||||||||||||  
 Db 66 GATFOVEVPGSOHTDSOKKAI 86

Search completed: October 24, 2002, 15:20:28  
 Job time : 34.7377 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:18:53 : Search time 12.3934 Seconds  
(without alignments)  
41.388 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATFQVEVPGSQHIDSOKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCROS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	102	3	US-08-952-337-5
2	106	100.0	102	3	US-08-952-337-6
3	106	100.0	103	2	US-08-472-171-2
4	106	100.0	103	2	US-08-894-526-2
5	106	100.0	103	2	US-09-013-047-2
6	106	100.0	103	4	US-09-374-597-2
7	106	100.0	103	4	US-09-191-852-21
8	106	100.0	103	5	PCF-US95-13376-21
9	106	100.0	123	3	US-08-952-337-1
10	106	100.0	123	3	US-08-952-337-2
11	106	100.0	124	2	US-08-747-410-2
12	101	95.3	93	2	US-08-292-968-26
13	101	95.3	93	2	US-08-467-974-26
14	101	95.3	93	2	US-08-467-536-26
15	101	95.3	93	2	US-08-467-976-26
16	101	95.3	93	4	US-09-082-514-26
17	101	95.3	371	4	US-08-829-026A-6
18	94	86.7	124	1	US-08-449-045C-4
19	94	86.7	124	2	US-08-435-605A-12
20	94	88.7	124	6	5223610-3
21	51	48.1	448	4	US-08-878-989-2
22	51	48.1	448	4	US-09-272-796-2
23	51	48.1	508	4	US-09-344-700-4
24	42	36.6	855	4	US-09-027-337-2
25	41.5	39.2	1090	4	US-09-346-237-5
26	40.5	38.2	856	3	US-08-709-784-2
27	40.5	38.2	862	2	US-08-209-521-23

28	40.5	38.2	862	2	US-08-209-521-30	Sequence 30, Appl
29	40.5	38.2	862	4	US-09-059-461-2	Sequence 2, Appl
30	40.5	38.2	862	4	US-08-961-810-133	Sequence 133, App
31	40.5	38.2	862	4	US-08-352-902D-133	Sequence 133, App
32	39	36.8	251	1	US-07-956-700B-94	Sequence 94, Appl
33	39	36.8	251	1	US-08-476-537-94	Sequence 94, Appl
34	39	36.8	251	1	US-08-485-607-94	Sequence 94, Appl
35	39	36.8	251	2	US-08-475-879-94	Sequence 94, Appl
36	39	36.8	346	2	US-08-602-359A-34	Sequence 34, Appl
37	39	36.8	384	4	US-09-025-580-35	Sequence 35, Appl
38	39	36.8	427	4	US-09-025-580-36	Sequence 36, Appl
39	39	36.8	459	6	5194375-6	Patent No. 5194375
40	39	36.8	693	4	US-09-564-805-234	Sequence 234, App
41	39	36.8	775	2	US-08-714-070A-1	Sequence 1, Appl
42	38	35.8	845	1	US-08-416-950-11	Sequence 11, Appl
43	38	35.8	845	2	US-08-469-830-11	Sequence 11, Appl
44	38	35.8	1073	4	US-09-541-782-6	Sequence 6, Appl
45	38	35.8	1810	5	PCF-US95-11684-4	Sequence 4, Appl

## ALIGNMENTS

```

RESULT 1
US-08-952-337-5
; Sequence 5, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; EARLIER FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-08-952-337-5

Query Match          100.0%; Score 106; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATFQVEVPGSQHIDSOKKAI 21
      44 GATFQVEVPGSQHIDSOKKAI 64

RESULT 2
US-08-952-337-6
; Sequence 6, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; EARLIER FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6

```

SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 102  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 100.0%; Score 106; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 44 GATFOVEVPGSOHIDSOKKAI 64

RESULT 3  
US-08-472-171-2  
Sequence 2, Application US/08472171  
Patent No. 5932714

GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,171  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393,334  
FILING DATE: 23-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-507 MIS:V9

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1153  
TELEFAX: 416-595-1155  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-171-2

US-08-472-171-2

Query Match 100.0%; Score 106; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 4

US-08-894-526-2  
Sequence 2, Application US/08894526  
Patent No. 5942418

GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM  
TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,526  
FILING DATE: 01-DEC-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-894-526-2

Query Match 100.0%; Score 106; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 5  
US-09-013-047-2  
Sequence 2, Application US/09013047  
Patent No. 5998168

GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.

TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, 6th Floor  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/013,047  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/472,171  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/393,334  
FILING DATE: 23-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-013-047-2

Query Match 100.0%; Score 106; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 6  
US-09-374-597-2  
Sequence 2, Application US/09374597  
Patent No. 6140082  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Vacoob, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression of Gene Products from  
TITLE OF INVENTION: Genetically Manipulated Strains of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, 6th Floor  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/374,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/393,334  
FILING DATE: FEBRUARY 23, 1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-964

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-374-597-2

Query Match 100.0%; Score 106; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 7  
US-09-191-852-21  
Sequence 21, Application US/09191852  
Patent No. 6194560  
GENERAL INFORMATION:  
APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/191,852  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
APPLICATION NUMBER: 08/817,906  
FILING DATE: 04-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: P01590US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-191-852-21

Query Match 100.0%; Score 106; DB 4; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 8  
PCT-US95-13376-21  
Sequence 21, Application PC/TUS9513376  
GENERAL INFORMATION:  
APPLICANT: The Texas A&M University System  
APPLICANT: 310 Wisenbaker  
APPLICANT: College Station, Texas 77843-3369  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jones, John W.  
REGISTRATION NUMBER: 31,380  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US95-13376-21  
Query Match 100.0%; Score 106; DB 5; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.5e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 9  
US-08-952-337-1  
Sequence 1, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/OD758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Vibrio cholerae

US-08-952-337-1  
Query Match 100.0%; Score 106; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
Db 65 GATFOVEVPGSQHIDSOKKAI 85

RESULT 10  
US-08-952-337-2  
Sequence 2, Application US/08952337  
Patent No. 6019973  
GENERAL INFORMATION:  
APPLICANT: Holmgren, Jan  
APPLICANT: Lebens, Michael R.  
TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
FILE REFERENCE: 3846/OD758  
CURRENT APPLICATION NUMBER: US/08/952,337  
CURRENT FILING DATE: 1998-01-05  
EARLIER APPLICATION NUMBER: PCT/SE96/00570  
EARLIER FILING DATE: 1996-05-02  
EARLIER APPLICATION NUMBER: SE 9501682-0  
EARLIER FILING DATE: 1995-05-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-08-952-337-2  
Query Match 100.0%; Score 106; DB 3; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.1e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
Db 65 GATFOVEVPGSQHIDSOKKAI 85

RESULT 11  
US-08-747-410-2  
Sequence 2, Application US/08747410  
Patent No. 5993820  
GENERAL INFORMATION:  
APPLICANT: BAGDASARIAN, Michael  
APPLICANT: IRELAND, James  
TITLE OF INVENTION: CHIMERIC LTB VACCINES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 5993820 West Center, 90 South Seventh St  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/747,410  
FILING DATE: 12-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Kettelberger, Denise M  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 11526.1-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/371-5268  
TELEFAX: 612/332-9081  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 124 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
US-08-747-410-2

Query Match 100.0%; Score 106; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 3.2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21  
Db 66 GATFOVEVPGSOHIDSOKKAI 86

RESULT 12  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 5856122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-292-968-26

Query Match 95.3%; Score 101; DB 2; Length 93;  
Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSOHIDSOKKAI 21  
Db 35 GATFOVEVPGSOHIDSOKKAI 55

RESULT 13  
US-08-467-974-26  
Sequence 26, Application US/08467974  
Patent No. 5965385  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,974  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/467,536  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-454 MIS:V9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-974-26

Query Match 95.3%; Score 101; DB 2; Length 93;  
 Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
 Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 14  
 US-08-467-536-26  
 ; Sequence 26, Application US/08467536  
 ; Patent No. 5977304  
 ; GENERAL INFORMATION:  
 ; APPLICANT: READ, Randy J.  
 ; APPLICANT: STEIN, Penelope E.  
 ; APPLICANT: COCKLE, Stephen A.  
 ; APPLICANT: OOMEN, Raymond P.  
 ; APPLICANT: LOOSMORE, Sheena  
 ; APPLICANT: KLEIN, Michel H.  
 ; APPLICANT: ARMSTRONG, Glen D.  
 ; APPLICANT: HAZES, Bart  
 ; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
 ; NUMBER OF SEQUENCES: 46  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & Mcburney  
 ; STREET: Suite 701, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/467,536  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/292,968  
 ; FILING DATE: 22-AUG-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/251,121  
 ; FILING DATE: 31-MAY-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/110,947  
 ; FILING DATE: 24-AUG-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: STEWART, Michael I.  
 ; REGISTRATION NUMBER: 24,973  
 ; REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 26:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 93 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-467-536-26

Query Match 95.3%; Score 101; DB 2; Length 93;  
 Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
 Db 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 15  
 US-08-467-976-26  
 ; Sequence 26, Application US/08467976  
 ; Patent No. 6018022  
 ; GENERAL INFORMATION:  
 ; APPLICANT: READ, Randy J.  
 ; APPLICANT: STEIN, Penelope E.  
 ; APPLICANT: COCKLE, Stephen A.  
 ; APPLICANT: OOMEN, Raymond P.  
 ; APPLICANT: LOOSMORE, Sheena  
 ; APPLICANT: KLEIN, Michel H.  
 ; APPLICANT: ARMSTRONG, Glen D.  
 ; APPLICANT: HAZES, Bart  
 ; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
 ; NUMBER OF SEQUENCES: 46  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & Mcburney  
 ; STREET: Suite 701, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/467,976  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/292,968  
 ; FILING DATE: 22-AUG-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/251,121  
 ; FILING DATE: 31-MAY-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/110,947  
 ; FILING DATE: 24-AUG-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: STEWART, Michael I.  
 ; REGISTRATION NUMBER: 24,973  
 ; REFERENCE/DOCKET NUMBER: 1038-453 MIS:vg  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 26:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 93 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-467-976-26

Query Match 95.3%; Score 101; DB 3; Length 93;  
 Best Local Similarity 95.2%; Pred. No. 1.7e-10;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
 Db 35 GETFOVEVPGSOHIDSOKKAI 55

Search completed: October 24, 2002, 15:23:59  
 Job time : 13.3934 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:20:33 : Search time 141.836 Seconds  
(without alignments)  
52.114 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATFQVEVPGSHIDSOKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database : Pending Patents\_AA\_Main:\*

```

1: /cgn2_6/ptodata/1/paa/pctus.COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097.COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098.COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099.COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100.COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101.COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US60.COMB.pep:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	21	US-09-786-648-4	Sequence 4, Appl
2	106	100.0	103	US-08-393-834-2	Sequence 2, Appl
3	106	100.0	103	US-08-782-832-15	Sequence 15, Appl
4	106	100.0	103	US-08-817-906-21	Sequence 21, Appl
5	106	100.0	103	US-09-836-433-14	Sequence 14, Appl
6	106	100.0	116	US-09-836-433-20	Sequence 20, Appl
7	106	100.0	119	US-09-836-433-22	Sequence 22, Appl

```

8 106 100.0 124 1 PCT-US99-30747-55 Sequence 55, Appl
9 106 100.0 124 18 US-09-470-124-55 Sequence 55, Appl
10 106 100.0 138 13 US-08-914-479-2 Sequence 2, Appl
11 106 100.0 138 13 US-08-914-479A-2 Sequence 2, Appl
12 106 100.0 313 21 US-09-756-983-15 Sequence 15, Appl
13 106 100.0 313 21 US-09-756-983-18 Sequence 18, Appl
14 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
15 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
16 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
17 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
18 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
19 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
20 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
21 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
22 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
23 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
24 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
25 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
26 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
27 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
28 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
29 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
30 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
31 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
32 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
33 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
34 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
35 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
36 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
37 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
38 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
39 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
40 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
41 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
42 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
43 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
44 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl
45 106 100.0 313 21 US-09-756-983-18 Sequence 22, Appl

```

## ALIGNMENTS

```

RESULT 1
US-09-786-648-4
; Sequence 4, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacc
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; PRIOR APPLICATION NUMBER: 2001-03-07
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 4
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic E. coli
; OTHER INFORMATION: human variant E. coli
US-09-786-648-4

```

Query Match 100.0%; Score 106; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATFOVEVPGSQHIDSOKKAI 21

RESULT 2  
US-08-393-334-2

Sequence 2, Application US/08393334  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & Mcburney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,334  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-393-334-2

Query Match 100.0%; Score 106; DB 7; Length 103;  
Best Local Similarity 100.0%; Pred. No. 9.8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21  
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 3  
US-08-782-832-15

Sequence 15, Application US/08782832  
GENERAL INFORMATION:  
APPLICANT: Arintzen, Charles J.  
APPLICANT: Mason, Hugh S.  
APPLICANT: Haq, Tariq A.  
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC  
TITLE OF INVENTION: BACTERIAL PROTEIN IN TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,832  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kettelberger, Denise M.  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 36170/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-782-832-15

Query Match 100.0%; Score 106; DB 11; Length 103;  
Best Local Similarity 100.0%; Pred. No. 9.8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSOKKAI 21  
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 4  
US-08-817-906-21

Sequence 21, Application US/08817906  
GENERAL INFORMATION:  
APPLICANT: Charles J. Arintzen, Hugh S. Mason, John D. Clements,  
APPLICANT: and Tariq A. Haq  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,906  
FILING DATE: 08/04/97  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid

TOPOLOGY: Linear  
US-08-817-906-21

Query Match 100.0%; Score 106; DB 12; Length 103;  
Best Local Similarity 100.0%; Pred. No. 9.8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 5  
US-09-836-433-14

; Sequence 14, Application US/09836433  
; GENERAL INFORMATION:  
; APPLICANT: Yuki, Yoshikazu  
; APPLICANT: Udaaka, Shigezo  
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
; FILE REFERENCE: Not Assigned  
; CURRENT APPLICATION NUMBER: US/09/836,433  
; CURRENT FILING DATE: 2001-04-16  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 103  
; TYPE: PRT  
; ORGANISM: Vibrio cholerae  
US-09-836-433-14

Query Match 100.0%; Score 106; DB 22; Length 103;  
Best Local Similarity 100.0%; Pred. No. 9.8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 6  
US-09-836-433-20

; Sequence 20, Application US/09836433  
; GENERAL INFORMATION:  
; APPLICANT: Yuki, Yoshikazu  
; APPLICANT: Udaaka, Shigezo  
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
; FILE REFERENCE: Not Assigned  
; CURRENT APPLICATION NUMBER: US/09/836,433  
; CURRENT FILING DATE: 2001-04-16  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20  
; LENGTH: 116  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-836-433-20

Query Match 100.0%; Score 106; DB 22; Length 116;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 7  
US-09-836-433-22

; Sequence 22, Application US/09836433  
; GENERAL INFORMATION:  
; APPLICANT: Yuki, Yoshikazu  
; APPLICANT: Udaaka, Shigezo  
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE

; FILE REFERENCE: Not Assigned  
; CURRENT APPLICATION NUMBER: US/09/836,433  
; CURRENT FILING DATE: 2001-04-16  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-836-433-22

Query Match 100.0%; Score 106; DB 22; Length 119;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 45 GATFOVEVPGSOHIDSOKKAI 65

RESULT 8  
PCT-US99-30747-55

; Sequence 55, Application PC/TUS9930747  
; GENERAL INFORMATION:  
; APPLICANT: Boyce Thompson Institute for Plant Research at Cor  
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In  
; FILE REFERENCE: 4868/85427  
; CURRENT APPLICATION NUMBER: PCT/US99/30747  
; CURRENT FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
PCT-US99-30747-55

Query Match 100.0%; Score 106; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSOHIDSOKKAI 86

RESULT 9  
US-09-470-124-55

; Sequence 55, Application US/09470124  
; GENERAL INFORMATION:  
; APPLICANT: Mason  
; APPLICANT: Amlizen  
; TITLE OF INVENTION: Orally Immunogenic Bacterial Enterotoxins Expressed In  
; FILE REFERENCE: 4868/84454  
; CURRENT APPLICATION NUMBER: US/09/470,124  
; CURRENT FILING DATE: 1999-12-24  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 55  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-470-124-55

Query Match 100.0%; Score 106; DB 18; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSOHIDSOKKAI 86

```

RESULT 10
US-08-914-479-2
; Sequence 2, Application US/08914479
; GENERAL INFORMATION:
; APPLICANT: Date, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 48112.404C2
; CURRENT APPLICATION NUMBER: US/08/914.479
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LT-B-M24 hybrid molecule
US-08-914-479-2

```

```

Query Match          100.0%; Score 106; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```

RESULT 11
US-08-914-479A-2
; Sequence 2, Application US/08914479A
; GENERAL INFORMATION:
; APPLICANT: Date, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 48112.404C2
; CURRENT APPLICATION NUMBER: US/08/914.479A
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LT-B-M24 hybrid molecule
US-08-914-479A-2

```

```

Query Match          100.0%; Score 106; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 66 GATFOVEVPGSQHIDSOKKAI 86

```

```

RESULT 12
US-09-756-983-15
; Sequence 15, Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Albanil, Salvatore

```

```

; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,
; TITLE OF INVENTION: CHARACTERIZATION AND MODULATION OF
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-15

```

```

Query Match          100.0%; Score 106; DB 21; Length 313;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 255 GATFOVEVPGSQHIDSOKKAI 275

```

```

RESULT 13
US-09-756-983-18
; Sequence 18, Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Albanil, Salvatore
; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,
; TITLE OF INVENTION: CHARACTERIZATION AND MODULATION OF
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-18

```

```

Query Match          100.0%; Score 106; DB 21; Length 351;
Best Local Similarity 100.0%; Pred. No. 4.6e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 293 GATFOVEVPGSQHIDSOKKAI 313

```

```

RESULT 14
US-09-756-983-22
; Sequence 22, Application US/09756983
; GENERAL INFORMATION:
; APPLICANT: Albanil, Salvatore
; TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,

```



```

; TITLE OF INVENTION: CHARACTERIZATION AND MODULATION OF
; TITLE OF INVENTION: ANTIGEN-SPECIFIC T CELLS
; FILE REFERENCE: 246/285-CIP
; CURRENT APPLICATION NUMBER: US/09/756,983
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/105,018
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-22

```

```

Query Match          100.0%; Score 106; DB 21; Length 364;
Best Local Similarity 100.0%; Pred. No. 4.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  GATFOVEVPGSOHIDSOKKAI 21
        |||
Db       306 GATFOVEVPGSOHIDSOKKAI 326

```

```

RESULT 15
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; TITLE OF INVENTION: Adjuvants
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45..65
; OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivable f
; OTHER INFORMATION: porcine E. coli
US-09-786-648-5

```

```

Query Match          95.3%; Score 101; DB 21; Length 21;
Best Local Similarity 95.2%; Pred. No. 9.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GATFOVEVPGSOHIDSOKKAI 21
        |
Db       1  GATFOVEVPGSOHIDSOKKAI 21

```

```

Search completed: October 24, 2002, 15:30:57
Job time : 142.836 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:21:08 ; Search time 47.5082 Seconds  
(without alignments)  
132.044 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEVPGSOHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 1105779

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PC/CT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	103	5	US-09-791-537-19387
2	106	100.0	103	5	US-09-791-537-38611
3	106	100.0	103	5	US-09-791-537-38639
4	106	100.0	103	5	US-09-791-537-74385
5	106	100.0	103	6	US-10-110-364-10
6	106	100.0	103	6	US-10-110-364-13
7	106	100.0	104	5	US-09-791-537-87980
8	106	100.0	104	5	US-09-791-537-99772
9	106	100.0	105	6	US-10-110-364-12
10	106	100.0	113	5	US-09-791-537-73608
11	106	100.0	123	6	US-10-110-364-17
12	106	100.0	124	5	US-09-791-537-40552
13	106	100.0	124	5	US-09-791-537-78640
14	106	100.0	124	5	US-09-791-537-92185
15	106	100.0	124	5	US-09-791-537-94644
16	106	100.0	124	5	US-09-791-537-103241
17	106	100.0	124	5	US-09-791-537-103660
18	106	100.0	124	6	US-10-110-364-6
19	106	100.0	124	6	US-10-110-364-11
20	106	100.0	124	6	US-10-110-364-19
21	106	100.0	124	6	US-10-110-364-21
22	106	100.0	124	6	US-10-110-364-21
23	106	100.0	125	6	US-10-110-364-15
24	106	100.0	129	5	US-09-791-537-131854
25	106	100.0	131	5	US-09-791-537-130348
26	106	100.0	138	6	US-10-141-627-2

27	101	95.3	103	6	US-10-110-364-22	Sequence 22, Appl
28	101	95.3	123	6	US-10-110-364-20	Sequence 20, Appl
29	101	95.3	124	5	US-09-791-537-28360	Sequence 28360, A
30	101	95.3	124	5	US-09-791-537-29489	Sequence 29489, A
31	101	95.3	124	5	US-09-791-537-60743	Sequence 60743, A
32	101	95.3	124	6	US-10-110-364-16	Sequence 16, Appl
33	101	95.3	124	6	US-10-110-364-18	Sequence 18, Appl
34	101	95.3	124	6	US-10-110-364-23	Sequence 23, Appl
35	100	94.3	103	5	US-09-791-537-68591	Sequence 68591, A
36	100	94.3	103	6	US-10-110-364-8	Sequence 8, Appl
37	100	94.3	124	5	US-09-791-537-33623	Sequence 33623, A
38	97	91.5	103	5	US-09-791-537-42610	Sequence 42610, A
39	97	91.5	103	5	US-10-110-364-7	Sequence 7, Appl
40	94	88.7	103	5	US-09-791-537-129309	Sequence 129309, A
41	94	88.7	103	6	US-10-110-364-5	Sequence 5, Appl
42	94	88.7	124	5	US-09-791-537-123948	Sequence 123948, A
43	94	88.7	124	6	US-10-110-364-4	Sequence 4, Appl
44	94	88.7	124	6	US-10-110-364-9	Sequence 9, Appl
45	94	88.7	382	1	PCT-US02-20978-1	Sequence 1, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-791-537-19387
; Sequence 19387, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19387
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1CT1D
US-09-791-537-19387

Query Match          100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB      45  GATFOVEVPGSOHIDSOKKAI 21
      1  |||
      2  |||
      3  |||
      4  |||
      5  |||
      6  |||
      7  |||
      8  |||
      9  |||
      10 |||
      11 |||
      12 |||
      13 |||
      14 |||
      15 |||
      16 |||
      17 |||
      18 |||
      19 |||
      20 |||
      21 |||
      22 |||
      23 |||
      24 |||
      25 |||
      26 |||
      27 |||
      28 |||
      29 |||
      30 |||
      31 |||
      32 |||
      33 |||
      34 |||
      35 |||
      36 |||
      37 |||
      38 |||
      39 |||
      40 |||
      41 |||
      42 |||
      43 |||
      44 |||
      45 |||

RESULT 2
US-09-791-537-38611
; Sequence 38611, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38611
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1CHPD
US-09-791-537-38611

```

```
Query Match          100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 3
US-09-791-537-38639
; Sequence 38639, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38639
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1CHOD
US-09-791-537-38639

Query Match          100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 4
US-09-791-537-74385
; Sequence 74385, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74385
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1FCBD
US-09-791-537-74385

Query Match          100.0%; Score 106; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 5
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio

APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511
US-10-110-364-10

Query Match          100.0%; Score 106; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 6
US-10-110-364-13
; Sequence 13, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: AB5 TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235
US-10-110-364-13

Query Match          100.0%; Score 106; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
|||||
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 7
US-09-791-537-87980
; Sequence 87980, Application US/09791537
; GENERAL INFORMATION:
```

```
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 87980
; LENGTH: 104
; TYPE: PRT
; ORGANISM: pdd 2CHBD
US-09-791-537-87980

Query Match          100.0%; Score 106; DB 5; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
   ||||||||||||||||
Db 46 GATFOVEVPGSQHIDSQKAI 66

RESULT 8
US-09-791-537-99772
; Sequence 99772, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 99772
; LENGTH: 104
; TYPE: PRT
; ORGANISM: pdd 3CHBD
US-09-791-537-99772

Query Match          100.0%; Score 106; DB 5; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
   ||||||||||||||||
Db 46 GATFOVEVPGSQHIDSQKAI 66

RESULT 9
US-10-110-364-12
; Sequence 12, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
```

```
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(105)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (Ogawa
; OTHER INFORMATION: 41).
US-10-110-364-12

Query Match          100.0%; Score 106; DB 6; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.4e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
   ||||||||||||||||
Db 46 GATFOVEVPGSQHIDSQKAI 66

RESULT 10
US-09-791-537-73608
; Sequence 73608, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 73608
; LENGTH: 113
; TYPE: PRT
; ORGANISM: pdd 1LTRD
US-09-791-537-73608

Query Match          100.0%; Score 106; DB 5; Length 113;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSQKAI 21
   ||||||||||||||||
Db 45 GATFOVEVPGSQHIDSQKAI 65

RESULT 11
US-10-110-364-17
; Sequence 17, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
```

OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.  
US-10-110-364-17

Query Match 100.0%; Score 106; DB 6; Length 123;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSQHIDSOKKAI 86

## RESULT 12

US-09-791-537-40552  
Sequence 40552, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 40552  
LENGTH: 124  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-09-791-537-40552

Query Match 100.0%; Score 106; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSQHIDSOKKAI 86

## RESULT 13

US-09-791-537-78640  
Sequence 78640, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 78640  
LENGTH: 124  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-09-791-537-78640

Query Match 100.0%; Score 106; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSQHIDSOKKAI 86

## RESULT 14

US-09-791-537-92185  
Sequence 92185, Application US/09791537

GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 92185  
LENGTH: 124  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-791-537-92185

Query Match 100.0%; Score 106; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSQHIDSOKKAI 86

## RESULT 15

US-09-791-537-94644  
Sequence 94644, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 94644  
LENGTH: 124  
TYPE: PRT  
ORGANISM: Vibrio cholerae  
US-09-791-537-94644

Query Match 100.0%; Score 106; DB 5; Length 124;  
Best Local Similarity 100.0%; Pred. No. 2.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 66 GATFOVEVPGSQHIDSOKKAI 86

Search completed: October 24, 2002, 15:33:19  
Job time: 47.5082 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:17:58 ; Search time 16.1803 Seconds  
(without alignments)  
124.712 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106  
Sequence: 1 GATPQVEVPGSGHIDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	95.3	124	1	OLECB
2	100	94.3	124	1	XVVCB
3	48	45.3	392	2	T04150
4	46	43.4	374	2	T19866
5	46	43.4	574	2	C86400
6	44.5	42.0	374	2	D81715
7	44.5	42.0	650	2	AG0733
8	44	41.5	91	2	C96580
9	44	41.5	255	2	A86457
10	44	41.5	1742	2	S76110
11	43	40.6	365	2	H69231
12	43	40.6	399	2	T46898
13	43	40.6	439	2	T49189
14	43	40.6	935	2	AE2469
15	42.5	40.1	376	1	S17246
16	42	39.6	216	2	F83962
17	42	39.6	299	2	JG0178
18	42	39.6	427	2	JC5694
19	42	39.6	787	2	G81692
20	42	39.6	864	2	T49574
21	41.5	39.2	500	2	JC4022
22	41.5	39.2	1090	2	S11823
23	41	38.7	91	2	F81109
24	41	38.7	128	1	UD00R
25	41	38.7	154	1	UD00R7
26	41	38.7	228	2	B34080
27	41	38.7	229	2	D27806
28	41	38.7	313	2	E88216
29	41	38.7	380	2	C34080

30	41	38.7	380	2	B34080	polyubiquitin 5 (c
31	41	38.7	381	2	A27806	polyubiquitin 5 (c
32	41	38.7	386	1	S66056	yaan protein - Bac
33	41	38.7	532	2	A34080	polyubiquitin 7 (c
34	41	38.7	550	2	T01770	hypothetical prote
35	41	38.7	2911	2	T20566	hypothetical prote
36	40.5	38.2	862	2	S47598	mult. protein homol
37	40	37.7	79	2	C95152	IS66 family elemen
38	40	37.7	79	2	H95167	IS66 family elemen
39	40	37.7	105	2	G69903	hypothetical prote
40	40	37.7	227	2	T32894	hypothetical prote
41	40	37.7	242	2	T34767	hypothetical prote
42	40	37.7	260	2	C96827	protein F20B17.2 (
43	40	37.7	330	2	C89848	peptide chain rele
44	40	37.7	355	2	T24938	hypothetical prote
45	40	37.7	367	2	T01751	gibberellin 20-oxi

## ALIGNMENTS

RESULT 1  
OLECB  
heat-labile enterotoxin chain B precursor - Escherichia coli  
C:Species: Escherichia coli  
C>Date: 29-Jun-1981 #sequence, revision 29-Jun-1981 #ext, change 18-Jun-1999  
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
R:Dallas, W.S.; Falkow, S.  
N:ature 288, 499-501, 1980  
A:Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat  
A:Reference number: A01820; MID:81074965  
A:Accession: A01820  
A:Molecule type: mRNA  
A:Residues: 1-124 <DAL>  
R:Yamamoto, T.; Gojobori, T.; Yokota, T.  
J. Bacteriol. 169, 1352-1357, 1987  
A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia  
A:Reference number: A26946; MID:87137303  
A:Accession: B26946  
A:Molecule type: DNA  
A:Residues: 1-27, 'E', '29-63, 'K', '65-124 <YAM>  
A:Cross-references: EMBL:M5363; NID:g148335; PIDN:AAA24792.1; PID:g148336  
R:Jeong, J.; Vinal, A.C.; Dallas, W.S.  
Infect. Immun. 48, 73-77, 1985  
A:Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
A:Reference number: I41194; MID:85156481  
A:Accession: I41194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5, 'F', '7-17, 'C', '19-24, 'S', '26-27, 'E', '29-33, 'H', '35-63, 'K', '65-66, 'A', '68-122  
A:Cross-references: GB:M17874; NID:g145830; PIDN:AAA98064.1; PID:g145831  
A:Experimental source: Plasmid ENT-R PCG86  
R:Ibrahim, I.; Gentz, R.  
J. Biol. Chem. 262, 10189-10194, 1987  
A:Title: A functional interaction between the signal peptide and the translation appa  
ratus.  
A:Reference number: I41287; MID:87280041  
A:Accession: I41287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-22 <RE2>  
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376  
R:Inoue, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
FEBS Microbiol. Lett. 108, 157-161, 1993  
A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
A:Reference number: I53542; MID:93252225  
A:Accession: I67644  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-17, 'C', '19, 'Y', '21-24, 'S', '26-27, 'E', '29-63, 'K', '65-66, 'A', '68-122, 'E', '124 <R  
A:Cross-references: GB:S60731; NID:g408994; PIDN:AAC60441.1; PID:g408996  
R:Tsuji, T.; Iida, T.; Honda, T.; Mawatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
Microb. Pathog. 2, 361-390, 1987

A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin iso1  
 A:Reference number: A61475; MUID:89180953  
 A:Accession: A61475  
 A:Molecule type: Protein  
 A:Residues: 22-24,'S',26-27,'E',29-63,'K',65-66,'A',68-95,'A',97-122,'E',124 <TSU>  
 A:Experimental source: strain 240-3  
 C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six B c  
 C:Function:  
 A:Description: the biological activity of the toxin is produced by the A chain, which ac  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status predicted

Query Match 95.3%; Score 101; DB 1; Length 124;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GATFOVEVPGSOHDSOKKAI 21  
 |||  
 Db 66 GETFOVEVPGSOHDSOKKAI 86

RESULT 2  
 XVCB  
 cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N169  
 N:Alternate names: enterotoxin beta chain  
 C:Species: Vibrio cholerae  
 C:Date: 24-Apr-1984 #sequence revision 01-Sep-2000 #text change 02-Feb-2001  
 C:Accession: S14624; S39238; S39241; H82196; JCI078; SI7666; PC1010; A05130; A01819; A36  
 R:Dams, E. de Wolf, M.; Dierick, W.  
 Submitted to the EMBL Data Library, March 1991  
 A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio cholera  
 A:Reference number: S14623  
 A:Accession: S14624  
 A:Molecule type: DNA  
 A:Residues: 1-124 <DAM>  
 A:Cross-references: EMBL:X56786; NID:948420; PIDN:CAAM1593.1; PID:948422  
 A:Experimental source: strain 2125  
 R:Lebens, M.; Holmgren, J.  
 Submitted to the EMBL Data Library, November 1993  
 A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera 01  
 A:Reference number: S39238  
 A:Accession: S39238  
 A:Molecule type: DNA  
 A:Residues: 1-124 <LEB>  
 A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAAM53973.1; PID:9433857  
 A:Accession: S39241  
 A:Molecule type: DNA  
 A:Residues: 1-124 <LEW>  
 A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAAM53976.1; PID:9433861  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406:477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833  
 A:Accession: H82196  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-124 <HEI>  
 A:Cross-references: GB:AE004224; GB:AE003852; NID:9655952; PIDN:AAE94613.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.  
 Chinese Biochem. J. 9, 395-399, 1993  
 A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.chole  
 A:Reference number: JCI078  
 A:Accession: JCI078  
 A:Molecule type: DNA  
 A:Residues: 1-20,'Q',22-31,'Q',33-38,'H',40-49,'G',51-67,'T',69-124 <SHI>  
 A:Experimental source: classical biotype strain 569B  
 R:Dams, E.; de Wolf, M.; Dierick, W.

Biochim. Biophys. Acta 1090, 139-141, 1991  
 A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic  
 A:Reference number: S17665; MUID:91355224  
 A:Accession: S17666  
 A:Molecule type: DNA  
 A:Residues: 1-38,'H',40-67,'T',69-124 <DA>  
 A:Cross-references: EMBL:X58785; NID:948868; PIDN:CAAM1591.1; PID:948890  
 R:Ma, Q.J.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
 Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
 A:Title: B subunit of cholera toxin produced in Escherichia coli.  
 A:Reference number: PC1010  
 A:Accession: PC1010  
 A:Molecule type: Protein  
 A:Residues: 22-38,'H',40-41 <MAQ>  
 R:Mekalanos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde,  
 Nature 306, 551-557, 1983  
 A:Reference number: A93320; MUID:84068199  
 A:Accession: A05130  
 A:Molecule type: DNA  
 A:Residues: 1-32,'S',34-74,'S',76-124 <MEK>  
 A:Cross-references: GB:X00171; NID:948347; PIDN:CAAM24996.1; PID:9758351  
 R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
 J. Biol. Chem. 252, 7257-7264, 1977  
 A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
 A:Reference number: A01819; MUID:78005537  
 A:Accession: A01819  
 A:Molecule type: Protein  
 A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <KUR>  
 R:Lat, C.Y.  
 J. Biol. Chem. 252, 7249-7256, 1977  
 A:Title: Determination of the primary structure of cholera toxin B subunit.  
 A:Reference number: A38033; MUID:78005536  
 A:Accession: A38033  
 A:Molecule type: Protein  
 A:Residues: 22-38,'H',40-42,'N',44-67,'T',69,'E',71-90,'N',92-124 <DAI>  
 A:Note: the difference at residue 70 may be due to denaturation during preparation  
 R:Nakashima, Y.; Napiorkowski, P.; Schaefer, D.E.; Konigsberg, W.H.  
 FEBS Lett. 68, 275-278, 1976  
 A:Title: Primary structure of the B subunit of cholera enterotoxin.  
 A:Reference number: A38034; MUID:77026365  
 A:Accession: A38034  
 A:Molecule type: Protein  
 A:Residues: 22-38,'H',40-67,'T',69,'E',71,'QS',74-75,'VE',78-86,'Q',88-99,'Q',101-103  
 R:Takao, T.; Watanabe, H.; Shimonishi, Y.  
 Eur. J. Biochem. 146, 503-508, 1985  
 A:Title: Facile identification of protein sequences by mass spectrometry.  
 A:Reference number: A21910; MUID:85126976  
 A:Accession: A21910  
 A:Molecule type: Protein  
 A:Residues: 22-38,'H',40-42,'N',44-67,'T',69-90,'N',92-124 <TAK>  
 A:Experimental source: biotype Inaba 569B  
 A:Note: Asp-65 was partially deaminated to Asp  
 C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
 C:Genetics:  
 A:Gene: VC1456  
 A:Map position: 1  
 C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha  
 ciate noncovalently with the subunit B, an aggregate of five beta chains  
 C:Function:  
 A:Description: involved in binding of the toxin to cell membranes  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin; toxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status experimental

Query Match 94.3%; Score 100; DB 1; Length 124;  
 Best Local Similarity 95.2%; Pred. No. 4.3e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GATFOVEVPGSOHDSOKKAI 21  
 |||  
 Db 66 GETFOVEVPGSOHDSOKKAI 86



## RESULT 3

RAD23 protein homolog - rice  
 T04150  
 C:Species: Oryza sativa (rice)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: T04150  
 R:Schultz, T.F.; Quatrano, R.S.  
 Plant Mol. Biol. 34, 557-562, 1997  
 A:Title: Characterization and expression of a rice RAD23 gene.  
 A:Reference number: 208695; MUID:97369378  
 A:Accession: T04150  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-392 <SCH>  
 A:Cross-references: EMBL:U63530; NID:g1488296; PIDN:AAB65841.1; PID:g1488297  
 A:Experimental source: cv. Nipponbare  
 C:Genetics:  
 A:Gene: RAD23  
 C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

## Query Match

Best Local Similarity 45.3%; Score 48; DB 2; Length 392;  
 Pred. 42.9%; Pred. No. 5.4;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSCHIDSOKKAI 21  
 DB 10 GSTFOIEVDSAQKADVKKRII 30

## RESULT 4

T19866  
 hypothetical protein C40H5.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T19866  
 R:White, S.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19189  
 A:Accession: T19866  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-374 <WIL>  
 A:Cross-references: EMBL:Z81482; PIDN:CAB03954.2; GSPDB:GN00028; CESP:C40H5.3  
 C:Genetics:  
 A:Gene: CESP:C40H5.3  
 A:Map position: X  
 A:Introns: 44/2; 95/1; 178/1; 234/3; 348/2  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C40H5.3

## Query Match

Best Local Similarity 43.4%; Score 46; DB 2; Length 374;  
 Pred. 42.1%; Pred. No. 11;  
 Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSCHIDSOKK 19  
 DB 230 GVEFHMFPGOPHADEOKE 248

## RESULT 5

C86400  
 hypothetical protein T17H3.2 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: C86400  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Reul, S.; White, O.; Alonso, Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Hutzar, L.  
 Nature 408, 815-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: C86400  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-574 <STO>  
 A:Cross-references: GB:AE005172; NID:95668763; PIDN:AAD45990.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

## Query Match

Best Local Similarity 43.4%; Score 46; DB 2; Length 574;  
 Pred. 58.8%; Pred. No. 17;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 TFOVEVPGSCHIDSOKK 19  
 DB 19 TFFVVDPSQHIEVKKK 35

## RESULT 6

DB1715  
 conserved hypothetical protein TC0328 [imported] - Chlamydia muridarum (strain N199)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: DB1715  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe, Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: DB1715  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-374 <TEM>  
 A:Cross-references: GB:AE002300; GB:AE002160; NID:g7190365; PIDN:AAF39192.1; PID:g719  
 A:Experimental source: strain N199 (MoPn)  
 C:Genetics:  
 A:Gene: TC0328

## Query Match

Best Local Similarity 42.0%; Score 44.5; DB 2; Length 374;  
 Pred. 48.0%; Pred. No. 20;  
 Matches 12; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

OY 2 ATFOVEVPGSCH-----IDSOKKAI 21  
 DB 321 ASPPDPLPGSQHKEKWDSTKRLAL 345

## RESULT 7

AG0733  
 probable bacteriophage protein STY2026 [imported] - Salmonella enterica subsp. enteri  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AG0733  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farr, S.; Moule, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AG0733  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-650 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05569.1; PID:g16503066; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY2026





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:42 ; Search time 8.26229 Seconds  
(without alignments)  
98.412 Million cell updates/sec

Title: US-09-786-648-4

Perfect score: 106

Sequence: 1 GATFOVEYPSQHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	124	1	ELBH_ECOLI
2	101	95.3	124	1	ELBP_ECOLI
3	100	94.3	124	1	CHTB_VIBCH
4	44	41.5	467	1	IRF6_HUMAN
5	44	41.5	467	1	IRP6_MOUSE
6	43	40.6	385	1	TGT_ZYMO
7	42.5	40.1	376	1	AROC_YEAST
8	42	39.6	855	1	ST14_HUMAN
9	41.5	39.2	500	1	GABT_HUMAN
10	41.5	39.2	1090	1	PUTA_KLEPN
11	41	38.7	76	1	UB1Q_DICDI
12	41	38.7	313	1	YP68_CAEEL
13	41	38.7	386	1	YVAN_BACSV
14	40.5	38.2	862	1	PMS2_HUMAN
15	40	37.7	288	1	HS74_CAEEL
16	40	37.7	309	1	CUT2_RAT
17	40	37.7	355	1	YRT1_CAEEL
18	40	37.7	655	1	YDH2_SCHPO
19	40	37.7	683	1	RPC_BPPHC
20	40	37.7	992	1	DUP1_RAT
21	40	37.7	2733	1	RRPB_CVMA5
22	39.5	37.3	560	1	INR1_SHEEP
23	39	36.8	112	1	Y13K_SSV1
24	39	36.8	263	1	YC43_ODOSI
25	39	36.8	456	1	SR54_THEAC
26	39	36.8	459	1	IL7R_MOUSE
27	39	36.8	491	1	CD5_RAT
28	39	36.8	661	1	HS7C_CAEEL
29	39	36.8	684	1	CC4_CANAL
30	39	36.8	704	1	PCCA_RAT
31	39	36.8	775	1	TH11_SCHPO
32	39	36.8	842	1	DPOL_HPVNA
33	39	36.8	1634	1	PK3B_HUMAN

34	39	36.8	2731	1	RRPB_CVMA5	P29982 murine coro
35	38.5	36.3	293	1	CHTA_CICAR	P36908 cicar ariet
36	38.5	36.3	344	1	LICH_PSESS	P25276 pseudomonas
37	38.5	36.3	529	1	TACY_LISMO	P13128 listeria mo
38	38.5	36.3	744	1	CATL_EMENI	P55305 emericella
39	38	35.8	76	1	UB1Q_NEUCR	P13117 neurospora
40	38	35.8	148	1	IF5A_AERPE	O9V653 aeropyrum p
41	38	35.8	169	1	Y459_MYCGE	O49436 mycoplasma
42	38	35.8	169	1	Y459_MYCPN	P75118 mycoplasma
43	38	35.8	205	1	HAP3_KLU0A	P40914 kluyveromyc
44	38	35.8	259	1	MOB2_YEAST	P43563 saccharomyc
45	38	35.8	359	1	THD1_SOLFV	P31212 solanum tub

## ALIGNMENTS

RESULT 1  
ID ELBH\_ECOLI STANDARD: PRT: 124 AA.  
AC P13811;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).  
GN ELTB OR LTBP.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia.  
NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H/4-114;  
RX MEDLINE=85156481; PubMed=3884513;  
RA Leong J., Vinal A.C., Dallas W.S.;  
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
RT cistrons from Escherichia coli of human and porcine origin.";  
RN Infect. Immun. 48:73-77(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H10407;  
RX MEDLINE=83114628; PubMed=6759877;  
RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
RT "Overlapping genes in the heat-labile enterotoxin operon originating  
RT from Escherichia coli human strain.";  
RN Mol. Gen. Genet. 188:356-359(1982).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE H10407;  
RX MEDLINE=93252225; PubMed=8486242;  
RA Inoue T., Tsuji T., Koto M., Imanura S., Miyama A.;  
RT "Amino acid sequence of heat-labile enterotoxin from chicken  
RT enterotoxigenic Escherichia coli is identical to that of human strain  
RT H 10407.";  
RN FEMS Microbiol. Lett. 108:157-161(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ETEC LT 87;  
RA Germani Y., Desperrier J.M.;  
RT Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=95349400; PubMed=7623669;  
RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappuoli R.;  
RT "Identification of errors among database sequence entries and  
RT comparison of correct amino acid sequences for the heat-labile  
RT enterotoxins of Escherichia coli and Vibrio cholerae.";  
RN Mol. Microbiol. 15:1165-1167(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
RX MEDLINE=99185101; PubMed=10085117;  
RA Matkovic-Calocovic D., Loregian A., D'Acunzio M.R., Battistutta R.,  
RA Tossi A., Palu G., Zanotti G.

```

RT      "Crystal structure of the B subunit of Escherichia coli heat-labile
RR      enterotoxin carrying peptides with anti-herpes simplex virus type 1
RL      activity.";
J. Biol. Chem. 274:8764-8769(1999).
CC      -I- EDUCATION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC      THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC      -I- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC      or send an email to: license@isb-slb.ch).
-----
DR      EMBL; M1874; AAA98064.1; -.
DR      EMBL; J01646; AAB02982.1; -.
DR      EMBL; S60731; AAC6044.1; -.
DR      EMBL; X83966; CA58800.1; -.
DR      PDB; 1LTR; 23-MAR-99.
DR      InterPro; IPR001835; Enterotoxin_B.
DR      Pfam; PF01376; Enterotoxin_B; 1.
DR      PRINTS; PR00772; ENTEROTOXINB.
DR      Prodom; PD012805; Enterotoxin.B; 1.
KW      Enterotoxin; Signal; 3D-structure.
FT      SIGNAL          1         21
FT      CHAIN           22        124   HEAT-LABILE ENTEROTOXIN B CHAIN.
FT      DISULFID        30        107
FT      VARIANT         6          6     F -> C (IN ISOLATE H10407) .
FT      VARIANTE        20         20     H -> Y (IN ISOLATE H10407) .
FT      VARIANTE        34         34     H -> R (IN ISOLATE H10407) .
SQ      SEQUENCE       124 AA;  14027 MW;  EBF7FC7B9D3BC47 CRC64;
Query Match               100.0%; Score 106; DB 1; Length 124;
Best Local Similarity    100.0%; Pred. No. 2,1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 GATFOVEVPGSOHIDSOKKAI 21
Db      66 GATFOVEVPGSOHIDSOKKAI 86
|||||
RESULT 2
ELBP_ECOLI STANDARD; PRT; 124 AA.
AC      P33890; P13768; P01557;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DE      15-JUL-1998 (Rel. 36, Last annotation update)
DE      Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTP-B).
OS      EltB OR LTPB.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ISOLATE P307;
RX      MEDLINE=81074965; PubMed=7003397;
RA      Dallas W.S., Falkow S.;
      "Amino acid sequence homology between cholera toxin and Escherichia
      coli heat-labile toxin.";
      Nature 288:499-501(1980).
[2]
RP      REVISIONS TO 28 AND 64.
RC      STRAIN-ISOLATE P307;
RX      MEDLINE=85156481; PubMed=3884513;
RA      Jeong J., Vinal A.C., Dallas W.S.;
      "Nucleotide sequence comparison between heat-labile toxin B-subunit
      ctions from Escherichia coli of human and porcine origin.";
      Infect. Immun. 48:73-77(1985).
[3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE PCG86;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojobori T., Yokota T.;
RT Evolutionary origin of pathogenic determinants in enterotoxigenic
RT Escherichia coli and Vibrio cholerae O1.";
RN J. Bacteriol. 169:1352-1357(1987).
[4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=87280041; PubMed=3301830;
RA Ibrahim I., Gentz R.;
RT "A functional interaction between the signal peptide and the
RT translation apparatus is detected by the use of a single point
RT mutation which blocks translocation across mammalian endoplasmic
RT reticulum.";
RN J. Biol. Chem. 262:10189-10194(1987).
[5]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX MEDLINE=93240541; PubMed=8478941;
RA Sliema T.K., van Zanten B.A.M., Deuter Z., Hol W.G.J.;
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a
RT close relative of cholera toxin.";
RN J. Mol. Biol. 230:890-918(1993).
[6]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91238966; PubMed=2034287;
RA Sliema T.K., Pronk S.E., Kalk K.H., Martina E.S., van Zanten B.A.M.,
RA Without B., Hol W.G.J.;
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin
RT from E. coli ";
RN Nature 351:371-377(1991).
[7]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=95349400; PubMed=7623669;
RA Domenighini M., Pizza M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae.";
RN Mol. Microbiol. 15:1165-1167(1995).
-1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
-1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL, M17873; AAA98065.1; -
CC EMBL, M15363; AAA24792.1; -
CC EMBL, M17101; AAA23973.1; -
CC PIR, A01820; OLECB.
CC PIR, B26946; OLECB.
CC PDB, 1LTA; 31-JAN-94.
CC PDB, 1LTB; 31-JAN-94.
CC PDB, 1LTG; 15-SEP-95.
CC PDB, 1LTI; 17-AUG-96.
CC PDB, 1LTS; 31-JAN-94.
CC PDB, 1LTT; 31-JAN-94.
CC PDB, 1LTT; 07-JUL-97.
CC PDB, 1LT4; 16-JUN-97.
CC PDB, 1LTS; 03-DEC-97.
CC PDB, 1LT6; 03-DEC-97.
CC PDB, 1HTT; 20-APR-95.
CC InterPro: IPR001835; Enterotoxin_B.
CC Pfam: PF01376; Enterotoxin_B. 1.
CC PRINTS: PR00772; ENTEROTOXINB.
CC ProDom: PD012805; Enterotoxin_B. 1.
CC Enterotoxin; Signal; 3D-structure.
CC
CC SIGNAL

```

FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT STRAND 36 43  
 FT STRAND 47 51  
 FT STRAND 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT HELIX 80 98  
 FT TURN 99 100  
 FT STRAND 103 109  
 FT STRAND 115 123  
 SO SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 95.3%; Score 101; DB 1; Length 124;  
 Best Local Similarity 95.2%; Pred. No. 1.4e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATGVEVPGSHIDSOKKAI 21  
 DB 66 GETGVEVPGSHIDSOKKAI 86

RESULT 3  
 CHTB\_VIBCH STANDARD; PRT: 124 AA.  
 ID CHTB\_VIBCH 09J002;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cholera enterotoxin, beta chain precursor.  
 GN CTXB OR TOXB OR VCI456.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84061784; PubMed=6315707;  
 RA Lockman H., Kaper J.B.;  
 RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio  
 cholerae enterotoxin.";  
 RL J. Biol. Chem. 258:13722-13726(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=EL TOR 2125;  
 RC MEDLINE=84068199; PubMed=6646234;  
 RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,  
 de Wilde M.;  
 RT "Cholera toxin genes: nucleotide sequence, deletion analysis and  
 vaccine development.";  
 RL Nature 306:551-557(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR 2125;  
 RA Dams E., de Wolf M., Dierick W.;  
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=4260B / SEROTYPE O139;  
 RX MEDLINE=94237453; PubMed=8181723;  
 RA Ledens M., Holmgren J.;  
 RT "Structure and arrangement of the cholera toxin genes in Vibrio  
 cholerae O139.";  
 RL FEMS Microbiol. Lett. 117:197-202(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1854 / O139-BENGAL;  
 RA Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,  
 Honda T.;  
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.

RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Drögl I., Sellers P.,  
 McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 RN [7]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005537; PubMed=903363;  
 RA Kurosky A., Markel D.E., Peterson J.W.;  
 RT "Covalent structure of the beta chain of cholera enterotoxin.";  
 RL J. Biol. Chem. 252:7257-7264(1977).  
 RN [8]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005536; PubMed=903362;  
 RA Lai C.-Y.;  
 RT "Determination of the primary structure of cholera toxin B subunit.";  
 RL J. Biol. Chem. 252:7249-7256(1977).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=94272319; PubMed=8003954;  
 RA Merritt E.A., Saffaty S., van den Akker F., L'Holr C., Martial J.A.,  
 Hol W.G.J.;  
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
 pentasaccharide.";  
 RL Protein Sci. 3:166-175(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=95387394; PubMed=7658472;  
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Otwinowski Z.,  
 Maulik P.R., Reed R.A., Shipley G.G.;  
 RT "The 2.4 Å crystal structure of cholera toxin B subunit pentamer:  
 choleraenoid.";  
 RL J. Mol. Biol. 251:550-562(1995).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RC STRAIN=OGAWA 41 / CLASSICAL BIOTYPE;  
 RX MEDLINE=97376625; PubMed=9232653;  
 RA Merritt E.A., Saffaty S., Jobling M.G., Chang T., Holmes R.K.,  
 Hirst T.R., Hol W.G.J.;  
 RT "Structural studies of receptor binding by cholera toxin mutants.";  
 RL Protein Sci. 6:1516-1528(1997).  
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
 BINDING TO CELL MEMBRANES.  
 CC -1- SUBUNIT: CONTAINS 3 KINS OF CHAINS. AN ALPHA AND A GAMMA CHAIN  
 (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
 DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
 6 BETA CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X00171; CAA24996.1; -  
 CC EMBL; R01170; AAA27573.1; -  
 CC EMBL; D30053; BAA06291.1; -  
 CC EMBL; X58786; CAA41593.1; -  
 CC EMBL; X76390; CAA53973.1; -  
 CC EMBL; X76391; CAA53976.1; -  
 CC EMBL; AE004224; AAF94613.1; -  
 CC PIR; A01819; XVCB.

```

DR PTR: A05130; A05130.
DR PTR: S14624; S14624.
DR PDB: 2CHB; 03-DEC-97.
DR PDB: 3CHB; 12-AUG-98.
DR PDB: 1CHP; 08-MAR-96.
DR PDB: 1CHO; 28-MAR-96.
DR PDB: 1ECB; 23-DEC-96.
DR PDB: 1XPB; 01-APR-97.
DR PDB: 1XTC; 01-AUG-96.
DR PDB: 1CT1; 15-OCT-97.
DR TIGR: VCI456; -.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B. 1.
DR PRINTS: PR00772; ENTEROTOXIN_B. 1.
DR PRODOM: PD012805; Enterotoxin_B; 1.
KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 124 CHOLERA ENTEROTOXIN, BETA CHAIN.
FT DISUFID 30 107
FT CONFLICT 33 33 Y -> S (IN REF. 2).
FT CONFLICT 39 39 Y -> H (IN REF. 7 AND 8).
FT CONFLICT 43 43 D -> N (IN REF. 7 AND 8).
FT CONFLICT 68 68 I -> T (IN REF. 7 AND 8).
FT CONFLICT 70 70 O -> E (IN REF. 8).
FT CONFLICT 75 75 G -> S (IN REF. 2).
FT CONFLICT 91 91 D -> N (IN REF. 7 AND 8).
FT HELIX 26 30
FT TURN 31 32
FT TURN 34 35
FT STRAND 36 44
FT STRAND 47 51
FT TURN 54 55
FT STRAND 58 62
FT TURN 64 65
FT STRAND 68 71
FT TURN 76 77
FT HELIX 80 99
FT TURN 100 100
FT STRAND 102 109
FT STRAND 115 123
SQ SEQUENCE 124 AA; 13957 MW; 9NA393E3EABE3EBF CRC64;

Query Match 94.3%; Score 100; DB 1; Length 124;
Best Local Similarity 95.2%; Pred. No. 2e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATFOVEVPGSQHIDSOKKAI 21
Db 66 GATFOVEVPGSQHIDSOKKAI 86

RESULT 4
IRF6_HUMAN
ID IRF6_HUMAN STANDARD; PRT; 467 AA.
AC O14896;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Interferon regulatory factor 6 (IRF-6).
GN IRF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RA Grossman A., Mitrucker H.W., Antonio L., Ozato K., Mak T.W.;
RU Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE FROM N.A.
RA Grafham D.;
RU Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

```

```

CC -!- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF027292; AAB8411.1; -.
DR EMBL: AL022398; CAA18545.1; -.
DR HSSP: P23906; 2IRF.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF. 1.
DR PRINTS: PR00267; INTERFEREFACT.
DR PRODOM: PD002355; IRF. 1.
DR SMART: SM00348; IRF. 1.
DR PROSITE: PS00601; IRF. 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.
SQ SEQUENCE 467 AA; 53129 MW; 7E28F5E0F5BA4053 CRC64;

Query Match 41.5%; Score 44; DB 1; Length 467;
Best Local Similarity 41.2%; Pred. No. 12;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 QVEVPGSQHIDSOKKAI 21
Db 273 QVKEPGPEHINKEOKL 289

RESULT 5
IRF6_MOUSE
ID IRF6_MOUSE STANDARD; PRT; 467 AA.
AC P97431;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Interferon regulatory factor 6 (IRF-6).
GN IRF6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Colon;
RA Grossman A., Mitrucker H.W., Antonio L., Mak T.W.;
RU Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [1]
RA -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U73029; AAB36714.1; -.
DR HSSP: P23906; 2IRF.
DR MGD: MGI:1859211; Irfg.
DR InterPro: IPR001346; IRF.
DR Pfam: PF00605; IRF. 1.
DR PRINTS: PR00267; INTERFEREFACT.
DR PRODOM: PD002355; IRF. 1.
DR SMART: SM00348; IRF. 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.
SQ SEQUENCE 467 AA; 53106 MW; 68CCAA90680FEDC8 CRC64;

```



Query Match 41.5%; Score 44; DB 1; Length 467;  
 Best Local Similarity 41.2%; Pred. No. 12;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 5 QVEVPGSOHIDSOBKRAI 21  
 Db 273 QVFPGEHITNEKRL 289

RESULT 6

TGT\_ZYMO STANDARD; PRT; 385 AA.  
 AC P28720; 060247;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Qyeine tRNA-ribosyltransferase (EC 2.4.2.29) (tRNA-guanine  
 transglycosylase) (Guanine insertion enzyme).  
 GN TGT.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
 OC Zymomonas.  
 OX NCBI\_TaxID=542;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-6, AND CHARACTERIZATION.  
 RA MEDLINE=95394847; PubMed=7665516;  
 RA Reuter K.K.H., Flicner R.;  
 RT "Sequence analysis and overexpression of the Zymomonas mobilis tgt  
 gene encoding tRNA-guanine transglycosylase: purification and  
 biochemical characterization of the enzyme.";  
 RL J. Bacteriol. 177:5284-5286(1995).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 31821 / ZM4 / CP4;  
 RX MEDLINE=92406015; PubMed=1526462;  
 RA Shark K.B., Conway T.;  
 RT "Cloning and molecular characterization of the DNA ligase gene (lig)  
 from Zymomonas mobilis.";  
 RL FEBS Microbiol. Lett. 75:19-26(1992).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).  
 RX MEDLINE=96256303; PubMed=8654383;  
 RA Romier C., Reuter K., Suck D., Flicner R.;  
 RT "Crystal structure of tRNA-guanine transglycosylase: RNA modification  
 by base exchange.";  
 RL EMBO J. 15:2850-2857(1996).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE=21103718; PubMed=1118905;  
 RA Graedler U., Gerber H.-D., Goodenough-Lashua D.M., Garcia G.A.,  
 RA Flicner R., Reuter K., Stubbs M.T., Kiebe G.;  
 RT "A new target for shigellosis: rational design and crystallographic  
 studies of inhibitors of tRNA-guanine transglycosylase.";  
 RL J. Mol. Biol. 306:455-467(2001).  
 CC -1- FUNCTION: Exchanges the guanine residue with 7-aminomethyl-7-  
 deazaguanine in tRNAs with G(U) anticodons (tRNA Asp, Asn, -His  
 and -Tyr). After this exchange, a cyclopentadiol moiety is  
 attached to the 7-aminomethyl group of 7-deazaguanine, resulting  
 in the hypermodified nucleoside queuosine (Q) (7-((4,5-cis-  
 dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine).  
 CC -1- CATALYTIC ACTIVITY: tRNA guanine + queine -> tRNA queine +  
 guanine.  
 CC -1- COFACTOR: BINDS AND REQUIRES ZINC FOR ACTIVITY. ALSO REQUIRES  
 MAGNESIUM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO THE QUEUINE tRNA-RIBOSYLTRANSFERASE FAMILY.  
 CC -1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY DUE  
 TO FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U33777; AAA27704.1; ALT\_INIT.  
 DR EMBL; U33777; AAA27705.1; ALT\_INIT.  
 DR EMBL; Z11910; -; NOT\_ANNOTATED\_CDS.  
 DR PDB; 1PUD; 07-JUL-97.  
 DR PDB; 1MKD; 07-JUL-97.  
 DR PDB; 1MKE; 07-JUL-97.  
 DR PDB; 1MKE; 07-JUL-97.  
 DR PDB; 1ENU; 19-APR-00.  
 DR InterPro: IPR002616; Que\_tRNA-ribosyltransf.  
 DR Pfam: PF01702; TGT; 1.  
 KW Queuosine biosynthesis; Transferase; Glycosyltransferase;  
 KW tRNA processing; Zinc; Magnesium; 3D-structure.  
 FT INTR\_MER 0 0  
 FT ACT\_SITE 102 102  
 FT ACT\_SITE 280 280  
 FT METAL 317 317 ZINC.  
 FT METAL 319 319 ZINC.  
 FT METAL 322 322 ZINC.  
 FT METAL 348 348 ZINC.  
 SQ SEQUENCE 385 AA; 42738 MW; F3D6FA270A0B23F3 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 385;  
 Best Local Similarity 38.1%; Pred. No. 14;  
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOBKRAI 21  
 Db 120 GATFKSHDGSRHMLSPERSI 140

RESULT 7

AROC YEAST STANDARD; PRT; 376 AA.  
 AC P28777;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Chorismate synthase (EC 4.6.1.4) (5-enolpyruvylshikimate-3-phosphate  
 dephosphorylase)  
 GN ARO2 OR YGL148W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=92114793; PubMed=1837329;  
 RA Jones D.G.L., Reusser U., Baus G.H.;  
 RT "Molecular cloning, characterization and analysis of the regulation  
 of the ARO2 gene, encoding chorismate synthase, of Saccharomyces  
 cerevisiae.";  
 RL Mol. Microbiol. 5:2143-2152(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1169;  
 RX MEDLINE=97197983; PubMed=9046099;  
 RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
 of chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
 RT TIP1, NFI1 genes and six new open reading frames.";  
 RL Yeast 13:177-182(1997).  
 CC -1- CATALYTIC ACTIVITY: 5-O-(1-carboxyvinyl)-3-phosphoshikimate ->  
 chorismate + phosphate.  
 CC -1- COFACTOR: REDUCED FLAVIN.  
 CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).



CC	HYPERREFLEXIA, LETHARGY, REFRACTORY SEIZURES, AND EEG
CC	ABNORMALITIES.
CC	-1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC	AMINOTRANSFERASES.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
DR	EMBL; L32961; AAA74449.1; -.
DR	HSSP; P80147; IGTX.
DR	MIM; 137150; -.
DR	InterPro; IPR000954; Aminotran_3.
DR	Pfam; PF00202; aminotran_3; 1.
DR	PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW	transferase; Aminotransferase; Pyridoxal phosphate;
KW	Neurotransmitter degradation; Mitochondrion; Transil peptide;
KW	Disease mutation.
FT	TRANSIT
FT	CHAIN
FT	BINDING
FT	VARIANT
FT	1 28
FT	29 500
FT	357 357
FT	220 220
FT	MITOCHONDRION.
FT	4-AMINOBUTYRATE AMINOTRANSFERASE.
FT	PYRIDOXAL PHOSPHATE.
FT	R -> K (IN GABA-AT DEFICIENCY; 258
FT	REDUCTION IN ACTIVITY).
FT	/FTID-VAR_008883.
FT	D -> H (IN REF. 2).
FT	V -> L (IN REF. 2).
FT	E -> G (IN REF. 2).
FT	K -> Q (IN REF. 2).
FT	W -> G (IN REF. 2).
FT	S -> A (IN REF. 2).
FT	G -> R (IN REF. 2).
FT	C -> G (IN REF. 2).
FT	L -> H (IN REF. 2).
SO	SEQUENCE 500 AA; 56557 MW; 41199085693F80AD CRC64;
QY	Query Match 39.2%; Score 41.5; DB 1; Length 500;
QY	Best Local Similarity 55.6%; Pred. No. 34;
QY	Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY	3 TFEVPEPGSQHIDSGKKA 20
QY	1:::1111111111
Db	17 TYRLVPGSRHT-SQAAA 33
RESULT 10	
PUL_A_KLEPN	STANDARD; PRT; 1090 AA.
AC	P07206;
DT	01-APR-1988 (Rel. 07, Created)
DT	01-APR-1990 (Rel. 14, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Pululanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6'-alpha-
DE	glucosidase) (Pululan 6-gluconohydrolase).
GN	PUL_A.
OS	Klebsiella pneumoniae.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Klebsiella.
OX	NCBI_TaxID=573;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=DNF 5023;
RC	MEDLINE=90205629; PubMed=2181242;
RA	Kornacker M.G., Pugsley A.P.;
RT	"Molecular characterization of pulA and its product, pululanase, a
RT	secreted enzyme of Klebsiella pneumoniae UNF5023.";
RL	Mol. Microbiol. 4:73-85(1990).
RN	[2]
RP	SEQUENCE OF 1-62 FROM N.A.
RC	MEDLINE=86033621; PubMed=3902792;
RA	

RA Chapon C., Raibaud O.;  
 RT "Structure of two divergent promoters located in front of the gene  
 RT encoding pullulanase in *Klebsiella pneumoniae* and positively  
 RT regulated by the malt product.";  
 RL J. Bacteriol. 164:639-645(1985).  
 RN [3]  
 RP SEQUENCE OF 944-1090 FROM N.A.  
 RX MEDLINE=89291709; PubMed=2661332;  
 RA D'Enfert C., Pugsley A.P.;  
 RT "Klebsiella pneumoniae pils gene encodes an outer membrane  
 RT lipoprotein required for pullulanase secretion.";  
 RL J. Bacteriol. 171:3673-3679(1989).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic  
 CC linkages in pullulan and in amylopectin and glycogen, and the  
 CC alpha-and beta-limit dextrans of amylopectin and glycogen.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 CC (probable).  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X52181; CAA36431.1; -;  
 DR EMBL: M12503; AAA25087.2; -;  
 DR EMBL: M29097; AAA61976.1; -;  
 DR PIR: A25025; A25025.  
 DR PIR: A32880; A32880.  
 DR PIR: S11823; S11823.  
 DR InterPro: IPR000461; Alpha.amylase.  
 DR InterPro: IPR004193; Isoamylase.N.  
 DR Pfam: PF00128; alpha-amylase; 1.  
 DR Pfam: PF02922; Isoamylase.N; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Hydroxylase; Glycosidase; Membrane; Lipoprotein; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 1 1090  
 FT LIPID 20 20 PULLULANASE.  
 FT ACT\_SITE 684 684 BY SIMILARITY.  
 FT ACT\_SITE 713 713 BY SIMILARITY.  
 FT ACT\_SITE 841 841 BY SIMILARITY.  
 FT CONFLICT 6 6 R -> C (IN REF. 2).  
 FT CONFLICT 10 10 V -> F (IN REF. 2).  
 FT CONFLICT 15 15 V -> I (IN REF. 2).  
 FT CONFLICT 23 23 G -> S (IN REF. 2).  
 FT CONFLICT 31 31 N -> S (IN REF. 2).  
 FT CONFLICT 34 34 T -> N (IN REF. 2).  
 FT CONFLICT 36 36 D -> DGNP (IN REF. 2).  
 FT CONFLICT 55 58 TAVE -> MATR (IN REF. 2).  
 SQ SEQUENCE 1090 AA; 118098 MW; 240AE7DEB3FF1BD6 CRC64;  
 Query Match 39.2%; Score 41.5; DB 1; Length 1090;  
 Best Local Similarity 38.5%; Pred. No. 78;  
 Matches 10; Conservative 4; Mismatches 7; Indels 5; Gaps 1;  
 QY 1 GATFOVEYPSQHTD-----SOKRAI 21  
 Db 312 GVTFVYVAPTAQYDVVVVYVSADKKVI 337  
 RESULT 11  
 UNIO\_DICDI STANDARD; PRT; 76 AA.  
 AC P08618;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-NOV-1991 (Rel. 20, Last annotation update)

DE Ubiquitin.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88152253; PubMed=2831095;  
 RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,  
 RA Gerisch G.;  
 RT "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-  
 RT terminal tail and identification of the protein using an anti-peptide  
 RT antibody.";  
 RL FEBS Lett. 229:273-278(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89352609; PubMed=2548604;  
 RA Omachi T., Giorda R., Shaw D.R., Ennis H.L.;  
 RT "Molecular organization of developmentally regulated Dictyostelium  
 RT discoideum ubiquitin cDNAs.";  
 RL Biochemistry 28:5226-5231(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87257921; PubMed=3037345;  
 RA Giorda R., Ennis H.L.;  
 RT "Structure of two developmentally regulated Dictyostelium discoideum  
 RT ubiquitin genes.";  
 RL Mol. Cell. Biol. 7:2097-2103(1987).  
 RN [4]  
 RP SEQUENCE OF 13-76 FROM N.A.  
 RC STRAIN=AX2;  
 RA Westphal M., Mueller-Taubenberger A., Noegel A., Gerisch G.;  
 RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
 RT Dictyostelium discoideum.";  
 RL FEBS Lett. 209:92-96(1986).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
 CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07210; CAA30183.1; ALT\_TERM.  
 DR EMBL: M19666; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19491; AAA33269.1; ALT\_TERM.  
 DR EMBL: M19492; AAA33270.1; ALT\_TERM.  
 DR EMBL: M23748; AAA33262.1; ALT\_TERM.  
 DR EMBL: M23749; AAA33263.1; ALT\_TERM.  
 DR EMBL: M23750; AAA33264.1; ALT\_TERM.  
 DR EMBL: M23751; AAA33265.1; ALT\_TERM.  
 DR EMBL: M23752; AAA33266.1; ALT\_TERM.  
 DR EMBL: M23753; AAA33267.1; ALT\_TERM.  
 DR EMBL: M23754; AAA33268.1; ALT\_TERM.  
 DR EMBL: X04702; CAA28408.1; ALT\_TERM.  
 DR PIR: S00357; UODOR7.  
 DR PIR: E34080; UODOR7.  
 DR PIR: A27806; A27806.  
 DR PIR: B27806; B27806.  
 DR PIR: A34080; A34080.  
 DR PIR: B34080; B34080.  
 DR PIR: C34080; C34080.  
 DR PIR: D34080; D34080.  
 DR HSSP: P02248; IOBT.  
 DR Dictydb; DD05001; ubqa.

```

DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; Ubiquitin.1.
DR PRINTS: PR00348; UBQUITIN.
DR SMART: SM00213; UBQ.1.
DR PROSITE: PS00299; UBQUITIN_1; 1.
DR PROSITE: PS0053; UBQUITIN_2; 1.
DR Nucleic acid protein; Polypeptide.
FT SITE 48 48
FT BINDING 76 76
FT VARIANT 11 11
SQ SEQUENCE 76 AA; 8538 MW; 6427383968EFA84 CRC64;

Query Match
Best Local Similarity 42.9%; Pred. No. 5.3;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 GATFOVEVPSQHIDSOKRAI 21
Db 10 GXTITLEVSGSDNIENYAKAI 30

RESULT 12
YF68_CAEEL STANDARD; PRT; 313 AA.
AC 009217;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.
GN B0495.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Kirsten J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -i SIMILARITY: SOME. TO YEAST YDL087C AND S.POMBE SPCC16A11.13.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U21317; AAA62527.1; -
DR WormPep: B0495.8; CE01766.
KW Hypothetical protein.
SQ SEQUENCE 313 AA; 36977 MW; 000D23237621BFED0 CRC64;

Query Match
Best Local Similarity 46.2%; Pred. No. 25;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 EYVPSQHIDSOKR 19
Db 11 QLMGSOHYDNKER 23

RESULT 13
YAAN_BACSU STANDARD; PRT; 386 AA.
AC P37535;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 43.8 kDa protein in XPAC-ABRB intergenic region.
GN YAAN.

```

```

OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE OF 1-191 FROM N.A.
RA Bookstein C., Edwards C.W., Hulet F.M.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D26185; BAA05262.1; -
DR EMBL: M96156; AAA22892.1; -
DR EMBL: Z99104; CAB11802.1; -
DR Subtilisin; BG100090; yaan.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 191 191
SQ SEQUENCE 386 AA; 43830 MW; D921F3A0F6845EE CRC64;

Query Match
Best Local Similarity 38.7%; Score 41; DB 1; Length 386;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEYVPSQHIDSOKRAI 21
Db 22 IETPGSEVAKAEQV 37

RESULT 14
PMS2_HUMAN STANDARD; PRT; 862 AA.
ID PMS2_HUMAN
AC P54278;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PMS1 protein homolog 2 (DNA mismatch repair protein PMS2).
GN PMS2 OR PMSL2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endometrial tumor;
RX MEDLINE=94352394; PubMed=8072530;
RA Nicolaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,
RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann S.R.,
RA Fraser C.M., Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R.,
RA Petersen G.M., de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT "Mutations of two PMS homologues in hereditary nonpolyposis colon
RT cancer."
RL Nature 371:75-80(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Bronner C.E., Baker S.M., Morrison P.T., Warren G., Smith L.G.,
RA Lescoe M.K., Kane M., Earlbino C., Lipford J., Lindblom A.,
RA Tannergaard P., Bollag R.J., Godwin A.R., Ward D.C.,
RA Nordenskiold M., Fishel R., Kolodner R.D., Liskay R.M.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:16:03 : Search time 26.1639 seconds  
(without alignments)  
138.851 Million cell updates/sec

Title: US-09-786-648-4  
Perfect score: 106  
Sequence: 1 GATFQVEVPGSOHIDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	124	2	Q57193
2	106	100.0	124	2	Q9RP15
3	106	100.0	124	2	Q56635
4	106	100.0	124	2	Q93V32
5	106	100.0	124	2	Q94M01
6	100	94.3	103	2	Q9R646
7	52	49.1	552	10	Q9LJ80
8	51	48.1	508	4	Q99987
9	49	46.2	428	10	Q9LGM2
10	48	44.3	392	10	Q40742
11	47	44.3	395	5	Q9NKM5
12	46.5	43.9	1166	3	Q9P4E0
13	46	43.4	293	5	Q9G876
14	46	43.4	374	5	Q9G3M9
15	46	43.4	574	10	Q9SXC9
16	44.5	42.0	374	16	Q9PKY2

17	44	41.5	91	10	Q9SYF4	Q9SYF4 arabidopsis
18	44	41.5	255	10	Q9C815	Q9C815 arabidopsis
19	44	41.5	257	10	Q9C875	Q9C875 arabidopsis
20	44	41.5	330	11	Q9QZL7	Q9QZL7 mus musculus
21	44	41.5	467	6	Q9N136	Q9N136 ovis aries
22	44	41.5	467	11	Q91V00	Q91V00 mus musculus
23	44	41.5	1742	16	Q55583	Q55583 synechocyst
24	43.5	41.0	565	10	Q22511	Q22511 vitis vinif
25	43	40.6	298	10	Q43098	Q43098 psophocarpu
26	43	40.6	356	5	Q9NFP90	Q9NFP90 leishmania
27	43	40.6	365	17	Q27066	Q27066 methanother
28	43	40.6	399	2	Q9P5L7	Q9P5L7 zymomonas m
29	43	40.6	439	10	Q9LY67	Q9LY67 arabidopsis
30	42	39.6	216	16	Q9K922	Q9K922 bacillus ha
31	42	39.6	298	10	Q9SXM5	Q9SXM5 glycine max
32	42	39.6	299	10	Q9S7G9	Q9S7G9 glycine max
33	42	39.6	427	13	Q42099	Q42099 cyprinus ca
34	42	39.6	697	5	Q965W6	Q965W6 caenorhabdl
35	42	39.6	787	16	Q9PKK3	Q9PKK3 chlamydia m
36	42	39.6	802	10	Q947W6	Q947W6 oryza sativ
37	42	39.6	849	16	Q98LD6	Q98LD6 rhizobium l
38	42	39.6	918	3	Q9P606	Q9P606 neurospora
39	42	39.6	1109	10	Q94HM6	Q94HM6 oryza sativ
40	42	39.6	1352	10	Q94EB4	Q94EB4 oryza sativ
41	42	39.6	2535	10	Q948C7	Q948C7 oryza sativ
42	41	38.7	91	16	Q9JRT4	Q9JRT4 neisseria m
43	41	38.7	246	5	Q9Y9U6	Q9Y9U6 drosophila m
44	41	38.7	413	2	Q46961	Q46961 erwinia chr
45	41	38.7	550	10	Q04636	Q04636 arabidopsis

## ALIGNMENTS

RESULT 1

Q57193 PRELIMINARY; PRT; 124 AA.

AC Q57193; 01-NOV-1996 (TREMREL. 01, Created)

DT 01-NOV-1996 (TREMREL. 01, Last sequence update)

DT 01-JUN-2001 (TREMREL. 17, Last annotation update)

DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).

GN CTXB.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL STRAIN 569B;

RX MEDLINE=91355224; PubMed=1883840;

RA Dams E., De Wolf M., Dietrick W.;

RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae

RL RT classical strain 569B.";

RN Blochim. Biophys. Acta 1090:139-141(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL BIOTYPE 569B;

RA Shi C., Cao C., Zhang J., Ma Q.;

RN Chin. Biochem. J. 9:395-399(1993).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CLASSICAL BIOTYPE 569B;

RA Xu L.;

RP Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; X58785; CAA41591.1; -

DR EMBL; U25679; AAC34728.1; -

DR EMBL; A00931; CAA00098.1; -

DR HSSP; P01556; ZCHR.

DR InterPro; IPR001835; Enterotoxin-B.

DR Pfam; PF01376; Enterotoxin\_B; 1.

DR PRINTS; PR00772; ENTEROTOXINB.

DR Prodom; PD012805; Enterotoxin\_B; 1.

KW Signal.

```

FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 124 CHOLERA TOXIN B PROTEIN (CTB).
SQ SEQUENCE 124 AA; 13919 MW; D6BF83FEF7924EA3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 106; DB 2; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKA1 21
Db 66 GATFOVEVPGSQHIDSQKKA1 86

RESULT 2
O9RP15 PRELIMINARY; PRT; 124 AA.
AC O9RP15;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CHOLERA ENTEROTOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KN1H002;
RA Shin H.J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RT cassette from Vibrio cholerae KN1H002 isolated in Korea.";
RL Misalimurhag Hoji 35:205-210(1999).
DR EMBL; AF175708; AAD51360.1; -.
DR HSSP; P01556; ZCHB.
DR InterPro; IPR001835; Enterotoxin_B.
DR Pfam; PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR PRODOM; PD012805; Enterotoxin_B; 1.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FEF793E589 CRC64;

Query Match
Best Local Similarity 100.0%; Score 106; DB 2; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKA1 21
Db 66 GATFOVEVPGSQHIDSQKKA1 86

RESULT 3
O56635 PRELIMINARY; PRT; 124 AA.
AC O56635;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CHOLERA TOXIN PRECURSOR.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Mawatani T., Albert M.J.,
RA Honda T.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae
RT strains 854 (O139-bengal) and S7 (O37) from two outbreaks.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30052; BAA06289.1; -.
DR HSSP; P01556; ZCHB.
DR InterPro; IPR001835; Enterotoxin_B.
DR Pfam; PF01376; Enterotoxin_B; 1.

```

```

DR PRINTS; PR00772; ENTEROTOXINB.
DR PRODOM; PD012805; Enterotoxin_B; 1.
RW Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;

Query Match
Best Local Similarity 100.0%; Score 106; DB 2; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKA1 21
Db 66 GATFOVEVPGSQHIDSQKKA1 86

RESULT 4
O93V32 PRELIMINARY; PRT; 124 AA.
AC O93V32;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.
GN LTH B SUBUNIT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RA Komase K.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RX MEDLINE=95091056; PubMed=7998417;
RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
RA Hattori N., Watanabe K., Suzuki Y., Negamine T., Aizawa C., Oya A.,
RA Kurata T.;
RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with
RT a trace amount of the holotoxin as an adjuvant for nasal influenza
RT vaccine.";
RL Vaccine 12:1083-1089(1994).
DR EMBL; AB011677; BAA25726.1; -.
SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 106; DB 2; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSQHIDSQKKA1 21
Db 66 GATFOVEVPGSQHIDSQKKA1 86

RESULT 5
O94M01 PRELIMINARY; PRT; 124 AA.
AC O94M01;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CTXB.
GN CTXB.
OS Vibrio phage CTX.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=141904;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharyya T., Nandy R.K., Nair G.B.;
RT "The entire core region of the ctx-phi (ctx-prophage) in VCE 232, an
RT environmental strain of V. cholerae.";

```



RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF414369; AAL09682.1; -  
 SQ SEQUENCE 124 AA; 13884 MW; 8DAC251591B16891 CRC64;

Query Match 100.0%; Score 106; DB 9; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2,6e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
 DB 66 GATFOVEVPGSOHIDSOKKAI 86

## RESULT 6

O9R646 PRELIMINARY; PRT; 103 AA.

AC O9R646; PRELIMINARY; PRT; 103 AA.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=66;  
 RN [1]  
 RP MEDLINE=95303036; PubMed=7783690;  
 RX "Nakashima K., Eguchi Y., Nakasone N.;  
 RT "Characterization of an enterotoxin produced by Vibrio cholerae  
 RT 0139."  
 RL Microbiol. Immunol. 39:87-94(1995).  
 DR HSSP: P01556; IYTC.  
 DR InterPro: IPR001835; Enterotoxin\_B.  
 DR Pfam: PF01376; Enterotoxin\_B; 1.  
 DR PRINTS: PR00772; ENTEROTOXINB.  
 DR PRODOM: PD012805; Enterotoxin\_B; 1.  
 SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;

Query Match 94.3%; Score 100; DB 2; Length 103;  
 Best Local Similarity 95.2%; Pred. No. 2.2e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKAI 21  
 DB 45 GATFOVEVPGSOHIDSOKKAI 65

## RESULT 7

O9LJB0 PRELIMINARY; PRT; 552 AA.

AC O9LJB0; PRELIMINARY; PRT; 552 AA.  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE EMB|CAB7996.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
 RT TAC and BAC clones."

RL DNA Res. 7:217-221(2000).  
 DR EMBL: AP000606; BAB01195.1; -  
 SQ SEQUENCE 552 AA; 63036 MW; 740E16CDEBB2447 CRC64;

Query Match 49.1%; Score 52; DB 10; Length 552;  
 Best Local Similarity 58.8%; Pred. No. 2.3;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 QVEVPGSOHIDSOKKAI 21  
 DB 147 QVEVPGSOHIDSOKKAI 163

## RESULT 8

O99987 PRELIMINARY; PRT; 508 AA.

AC O99987; PRELIMINARY; PRT; 508 AA.  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE VRK2.  
 OS VRK2.  
 GN VRK2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE=98008921; PubMed=9344656;  
 RA Nezu J., Oku A., Jones M.H., Shlman M.;  
 RT "Identification of two novel human putative serine/threonine kinases,  
 RT VRK1 and VR K2, with structural similarity to Vaccinia virus B1R  
 RT kinase."  
 RL Genomics 45:327-331(1997).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AB000450; BAA19109.1; -  
 DR HSSP: 006486; 1CKT.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 2.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 508 AA; 58126 MW; 157FBF6F48511AF4 CRC64;

Query Match 48.1%; Score 51; DB 4; Length 508;  
 Best Local Similarity 50.0%; Pred. No. 3.1;  
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 GATFOVEVPGSOHIDSOKKA 20  
 DB 329 GOSINVTSPNSOKVDSOKAA 348

## RESULT 9

O9LGM2 PRELIMINARY; PRT; 428 AA.

AC O9LGM2; PRELIMINARY; PRT; 428 AA.  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ESTS A0056822(S20908).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

```

RT  clone:PO041E11.";
RN  Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RC  [2]
RP  SEQUENCE FROM N.A.
RA  STRAIN=CV, NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
RT  clone:PO433F09.";
RL  Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AP002521; BAA96762.1; -
DR  HSSP; AP002539; BAB08201.1; -
DR  InterPro: IPR001552; Acyl-CoA_dh..
DR  Pfam; PF00441; Acyl-CoA_dh..1.
DR  Pfam; PF02770; Acyl-CoA_dh..1.
DR  Pfam; PF02771; Acyl-CoA_dh..1.
DR  PROSITE; PS00073; ACYL_COA_DH_2; UNKNOWN_1.
SQ  SEQUENCE 428 AA; 46132 MW; 8D34E3698A86367 CRC64;

Query Match 46.2%; Score 49; DB 10; Length 428;
Best Local Similarity 52.6%; Pred. No. 5.6;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TFEVPGSGHIDSOKKAI 21
Db 183 TTATKVPGGHIDGKKMI 201

RESULT 10
ID 040742 PRELIMINARY; PRT; 392 AA.
AC 040742;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE OSAD3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_Taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA MEDLINE=97369378; PubMed=9225866;
RA Schultz T.F., Quatrano R.S.;
RT "Characterization and expression of a rice RAD23 gene.";
RL Plant Mol. Biol. 34:557-562(1997).
DR EMBL; U63530; AB65841.1; -
DR HSSP; P54725; IDV0.
DR InterPro: IPR000449; UBA.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBA; 1.
DR PROSITE; PS50053; UBQUITIN_2; 1.
SQ SEQUENCE 392 AA; 41754 MW; BD6E08574CCTCAGB CRC64;

Query Match 45.3%; Score 48; DB 10; Length 392;
Best Local Similarity 42.9%; Pred. NO. 7.6;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Oy 1 GATFVEVPGSGHIDSOKKAI 21
Db 10 GSTFQIEVDSKQVADVKRII 30

RESULT 11
ID 09NKD5 PRELIMINARY; PRT; 395 AA.
AC 09NKD5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)

```

```

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DR HYPOTHETICAL 43.6 KDA PROTEIN.
GN BG:DS01514.3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=94403001; PubMed=10471707;
RA Ashburner M., Mista S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Ceiniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Ceiniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA Butenhoif C., Champe M., Chavez G., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomosti M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Svistskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003408; AAF44847.1; -
DR Flybase; FBgn0028907; BG:DS01514.3.
KW Hypothetical protein.
SQ SEQUENCE 395 AA; 43561 MW; AE4F1CC4ADD3DA73 CRC64;

Query Match 44.3%; Score 47; DB 5; Length 395;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 5 QVEVPGSGHIDSOKKA 20
Db 329 RVSVPGSTHIDADANA 344

RESULT 12
ID 09P4E0 PRELIMINARY; PRT; 1166 AA.
AC 09P4E0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PHEROMONE-RESPONSIVE MAPKK KINASE UBQ4.
DE UBQ4.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_Taxid=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1/2;
RX MEDLINE=20331594; PubMed=10875339;
RA Andrews D.L., Egan J.D., Mayorga M.E., Gold S.E.;
RT "The Ustilago maydis ubc4 and ubc5 genes encode members of a MAP
RT kinase cascade required for filamentous growth.";
RL MOL. Plant Microbe Interact. 13:781-786(2000).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF197562; AAF86841.1; -
DR HSSP; P24941; 1B38.
DR InterPro: IPR000719; Euk_Kinase.
DR InterPro: IPR000159; RA.

```

DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1166 AA; 123298 MW; A89E0019ABD4172 CRC64;

Query Match 43.9%; Score 46.5; DB 3; Length 1166;  
 Best Local Similarity 47.6%; Pred. No. 47;  
 Matches 10; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

OY 2 ATFOVEVP-GSOHIDSOKKAI 21  
 DB 729 AVKQVELPESGSHDQKKGM 749

RESULT 13

ID 049876 PRELIMINARY; PRT; 293 AA.  
 AC 049876.  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CLASS.III CHITINASE PRECURSOR (EC 3.2.1.14).  
 GN LUPAL.  
 OS Lupinus albus (White lupine).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.  
 OX NCBI\_TaxID=3870;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CV. ULTRA;  
 RA Regalado A.P., Vidal S., Neves A., Ricardo C.P.P.,  
 RA Rodrigues-Pousada C.;  
 RT "Constitutive and stress-induced expression of a class III chitinase  
 from Lupinus albus.";  
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 RT -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 HYDROLASES).  
 CC EMBL: Y16415; CAA76203.1; -.  
 DR HSSP: P23472; 2HW.  
 DR InterPro: IPR001579; Chitinase\_2.  
 DR Pfam: PF00192; chitinase\_2; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 KW Glycosidase; Hydrolase; Signal.  
 FT SIGNAL 1  
 FT CHAIN 28 293 POTENTIAL.  
 FT SEQUENCE 293 AA; 31128 MW; DB1B21728F657F2F CRC64;

Query Match 43.4%; Score 46; DB 10; Length 293;  
 Best Local Similarity 38.1%; Pred. No. 12;  
 Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATFOVEVP-GSOHIDSOKKAI 21  
 DB 142 GIDFIEAGCAOHYDELARAL 162

RESULT 14

ID 0903M9 PRELIMINARY; PRT; 374 AA.  
 AC 0903M9.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE C40H5.3 PROTEIN.  
 GN C40H5.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RP [1]  
 RC SEQUENCE FROM N.A.

RA White S.;  
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RT none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology.";  
 RL Science 283:2012-2018(1998).

DR EMBL: Z81482; CAB03954.2; -.  
 SQ SEQUENCE 374 AA; 42198 MW; 67D202886D6A7824 CRC64;

Query Match 43.4%; Score 46; DB 5; Length 374;  
 Best Local Similarity 42.1%; Pred. No. 16;  
 Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 GATFOVEVP-GSOHIDSOKK 19  
 DB 230 GVEFHIFPGQPHADEQKE 248

RESULT 15

ID 09SKC9 PRELIMINARY; PRT; 574 AA.  
 AC 09SKC9.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE T17H3.2 PROTEIN (AT1G27520/T17H3\_2).  
 GN T17H3.2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CV. COLUMBIA;  
 RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
 RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,  
 RA Chin C., Hong B., Choi E., Chlou J., Alcafi H., Araujo R., Brooks S.,  
 RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
 RA Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence.";  
 RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carrinchi P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cdna clones.";  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC005916; AAD45990.1; -.  
 DR EMBL: AY056382; AAL08238.1; -.  
 DR HSSP: P32906; 1DL2.  
 DR InterPro: IPR001382; Glyco\_hydro\_47.  
 DR Pfam: PF01532; Glyco\_hydro\_47; 1.  
 DR PRINTS: PR00747; GLYHDR47.  
 DR PRODOM: PD003239; Glyco\_hydro\_47; 1.  
 SQ SEQUENCE 574 AA; 65707 MW; 4ACC456DE487EA93 CRC64;

Query Match 43.4%; Score 46; DB 10; Length 574;  
 Best Local Similarity 58.8%; Pred. No. 26;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 3 TFOVEVPGSOHDSOKK 19  
| | | | | : | |  
Db 19 TFFVVDPSOHIYVKKK 35

Search completed: October 24, 2002, 15:22:28  
Job time : 29.4139 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:22:33 : Search time 32.7049 Seconds  
(without alignments)  
71.321 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107  
Sequence: 1 GETPQVEVPGSGHIDSKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 469622

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	21	AAV87463	E. coli heat labil
2	107	100.0	93	AAV72545	ADP-ribosylating t
3	107	100.0	93	AAV41816	Escherichia coli v
4	107	100.0	93	AAW95226	E. coli heat-labill
5	107	100.0	93	AAV68365	Heat labile toxin
6	107	100.0	93	AAV66239	E. coli verotoxin-1
7	101	94.4	21	AAV87462	Cholera toxin B/en
8	94	87.9	41	AAV50439	Network polymer wh
9	88	82.2	46	AAV50436	Network polymer wh
10	87.5	81.8	47	AAV30600	Sequence of amino
11	80	74.8	26	AAV30265	Sequence of amino

12	76	71.0	15	10	AAV93498	CRP3 epitope of th
13	76	71.0	15	16	AAV85125	Cholera toxin B an
14	76	71.0	23	16	AAV76748	Residues 50-64 of
15	62	57.9	12	21	AAV87461	Cholera toxin B/en
16	61	57.0	15	21	AAV15525	Cholera toxin B su
17	61	57.0	15	21	AAV15526	Bovine rotavirus V
18	59	55.1	15	22	AAV74320	Cholera toxin B su
19	43	40.2	8	21	AAV87464	Cholera toxin B/en
20	42	39.3	51	21	AAV02635	Human secreted pro
21	39	36.4	7	21	AAV87460	Cholera toxin B/en
22	39	36.4	75	18	AAV01700	Yeast ubiquitin.
23	39	36.4	75	19	AAV46380	Yeast ubiquitin le
24	39	36.4	75	20	AAV97614	Yeast ubiquitin am
25	39	36.4	76	16	AAV85462	Human ubiquitin.
26	39	36.4	76	19	AAV41801	Ubiquitin componen
27	39	36.4	76	21	AAV03161	Yeast ubiquitin.
28	38	35.5	48	22	AAV012358	Human polyubiquitin
29	38	35.5	51	22	AAV43738	Peptide #11244 enc
30	38	35.5	51	22	AAV26681	Protein #8680 enc
31	38	35.5	51	22	AAV64700	Human brain expres
32	38	35.5	51	22	AAV7476	Human bone marrow
33	38	35.5	51	22	AAV21398	Peptide #7832 enco
34	38	35.5	51	22	AAV37636	Peptide #11673 enc
35	38	35.5	57	22	AAV46087	Propionibacterium
36	38	35.5	69	21	AAV32702	Zea mays protein f
37	38	35.5	69	21	AAV40689	Zea mays protein f
38	38	35.5	71	21	AAV33547	Arabidopsis thalia
39	38	35.5	76	19	AAV60701	Ubiquitin encoded
40	38	35.5	76	21	AAV90254	Ubiquitin monomer
41	38	35.5	77	21	AAV06889	Arabidopsis thalia
42	38	35.5	77	21	AAV13297	Arabidopsis thalia
43	38	35.5	77	21	AAV15994	Arabidopsis thalia
44	38	35.5	77	21	AAV18728	Zea mays protein f
45	38	35.5	77	21	AAV42341	Arabidopsis thalia

## ALIGNMENTS

RESULT 1	AAV87463	standard; peptide; 21 AA.
ID	AAV87463;	
AC	AAV87463;	
XX		
DT	03-JUL-2000	(first entry)
XX		
DE	E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.	
XX		
KW	Heat labile enterotoxin subunit B; EtxB;	
KW	beta-4-alpha-2 loop; GM-1 ganglioside receptor; Immunomodulation;	
KW	adjuvant; immune disorder; diarrhoea.	
XX		
OS	Escherichia coli.	
XX		
PN	MO200014114-A1.	
PD	16-MAR-2000.	
XX		
PF	07-SEP-1999;	99NO-GB02970.
XX		
PR	07-SEP-1998;	98GB-0019484.
XX		
PA	(UYBR-) UNTV BRISTOL.	
XX		
PI	Williams NA, Hirst TR;	
XX		
DR	WPI; 2000-256943/22.	
XX		
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as	
PT	immunomodulators and for treating diarrhea and which do not bind the	
PT	glycolipid receptor GM-1 -	
XX		

PS Disclosure; Page 15; 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.

XX Sequence 21 AA;

SO Query Match 100.0%; Score 107; DB 21; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQSHDSOKKAI 21  
 Db 1 GETFOVEVPSQSHDSOKKAI 21

RESULT 2  
 AAR72545  
 ID AAR72545 standard; peptide; 93 AA.  
 AC AAR72545;  
 XX 28-NOV-1995 (first entry)  
 DT  
 XX ADP-ribosylating toxin (verotoxin-1 B-subunit).  
 DE  
 XX ADP-ribosylating toxin; pertussis holotoxin; B-subunit;  
 KW active site; E. coli heat labile toxin; verotoxin-1;  
 KM Bordetella pertussis vaccines.  
 XX  
 OS Bacteria sp.  
 XX  
 PI EP646599-A.  
 PN  
 XX 05-APR-1995.  
 PD  
 XX 23-AUG-1994; 94EP-0306219.  
 PF  
 XX 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 (UYAL-) UNIV ALBERTA.  
 XX  
 PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 PI Oomen R, Read RJ, Stein PE;  
 XX WPI; 1995-132623/18.  
 DR  
 XX New modified forms of pertussis holotoxin - developed using  
 PT crystalline forms of pertussis holotoxin and its complexes with  
 PT other molecules  
 XX  
 PS Disclosure; Fig 5; 54pp; English.

XX AAR72540-R72545 are structurally equivalent B-subunits from three  
 CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat  
 CC labile toxin (LT), and verotoxin-1 (VT). The structural  
 CC information obd. from these comparisons was used to identify  
 CC sites which contribute to PT's biological activity. By modifying  
 CC these sites the claimed PT mutants of the invention were produced,  
 CC they can be used in the development of vaccines against Bordetella  
 CC pertussis infection.

XX Sequence 93 AA;

SO Query Match 100.0%; Score 107; DB 16; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPSQSHDSOKKAI 21  
 Db 35 GETFOVEVPSQSHDSOKKAI 55

RESULT 3  
 AAY41816  
 ID AAY41816 standard; peptide; 93 AA.  
 AC AAY41816;  
 XX 08-DEC-1999 (first entry)  
 DT  
 XX Escherichia coli verotoxin-1 B-subunit.  
 DE  
 XX ADP-ribosylating toxin; PT: crystalline pertussis holotoxin; ETA; DT;  
 KW three-dimensional structure; LT; immunoprotective; infection.  
 XX  
 OS Escherichia coli.  
 XX  
 PN US5965385-A.  
 XX  
 PD 12-OCT-1999.  
 XX  
 PF 06-JUN-1995; 95US-0467974.  
 XX  
 PR 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 (UYAL-) UNIV ALBERTA.  
 XX  
 PI Read RJ, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Oomen RP;  
 XX WPI; 1999-579908/49.  
 DR  
 XX New method for producing modified pertussis holotoxin -  
 XX  
 PS Example 3; Fig 5; 41pp; English.

XX A method has been developed of producing a modified pertussis holotoxin,  
 CC involving analysis of the 3-dimensional form of the crystalline  
 CC holotoxin. The pertussis holotoxin modification process comprises:  
 CC (1) identification of at least one amino acid (aa) residue of the  
 CC holotoxin for modification by analysing the 3-dimensional form of the  
 CC crystalline holotoxin, in relation to known information of the protein  
 CC structure and function; (2) effecting mutagenesis (by removing or  
 CC replacing a nucleotide sequence encoding at least one (aa) of a tox  
 CC operon, and (3) expressing mutant tox box in a Bordetella organism to  
 CC produce the modified holotoxin. This method is used for modifying  
 CC pertussis holotoxin, by studying its 3-dimensional crystalline  
 CC structure. Modifying the holotoxin, alters its biological properties.  
 CC By analysing the 3-dimensional crystalline structure of the pertussis  
 CC holotoxin, functional (aa) which affect biological properties of the  
 CC pertussis holotoxin can be identified. This can be used to predict (aa)



QY 1 GETFOVEVPGSOHIDSOKKAI 21  
 DB 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 6  
 AAB66239  
 ID AAB66239 standard; Protein: 93 AA.  
 AC AAB66239;  
 DT 03-APR-2001 (first entry)  
 DE E coli verotoxin-1 B subunit SEQ ID NO: 26.  
 XX Pertussis toxin; crystal structure; whooping cough; biological activity;  
 KM lymphocytosis-promoting factor; histamine-sensitising factor;  
 KM 1alet-activating protein.  
 XX Escherichia coli.  
 OS  
 XX US6168928-B1.  
 PN  
 XX 02-JAN-2001.  
 PD  
 XX 21-MAY-1998; 98US-0082514.  
 PF  
 XX 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNNUGHT LAB LTD.  
 PI Read RJ, Cockle SA, Comen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Stein PE;  
 XX WPI; 2001-122260/13.  
 DR  
 XX Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site -  
 XX  
 PS Example 3; Fig 5; 41pp; English.  
 XX  
 CC The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.  
 CC  
 XX  
 SQ Sequence 93 AA;  
 QY Query Match 100.0%; Score 107; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 DB 35 GETFOVEVPGSOHIDSOKKAI 55

RESULT 7  
 AAB7462  
 ID AAB7462 standard; peptide: 21 AA.  
 AC AAB7462;  
 DT 03-JUL-2000 (first entry)

XX Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.  
 DE  
 XX Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
 KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
 KW adjuvant; immune disorder; diarrhoea.  
 XX Vibrio cholerae.  
 OS Escherichia coli.  
 XX  
 PN WO200014114-A1.  
 XX  
 PD 16-MAR-2000.  
 PF  
 XX 07-SEP-1999; 99WO-GB02970.  
 PR 07-SEP-1998; 98GB-0019484.  
 XX  
 PA (UYBR-) UNIV BRISTOL.  
 PI Williams NA, Hirst TR;  
 XX WPI; 2000-256943/22.  
 DR  
 XX Derivatives of Escherichia coli heat labile enterotoxins useful as  
 PT immunomodulators and for treating diarrhea and which do not bind the  
 PT glycolipid receptor GM-1 -  
 XX  
 PS Disclosure; Page 15; 62pp; English.  
 XX  
 CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAB7460-Y87463  
 CC represent preferred peptides of the invention, AAB7460 being  
 CC particularly preferred.  
 CC  
 XX  
 SQ Sequence 21 AA;  
 QY Query Match 94.4%; Score 101; DB 21; Length 21;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-10;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 DB 1 GATFOVEVPGSOHIDSOKKAI 21

RESULT 8  
 AAP50439  
 ID AAP50439 standard; protein: 41 AA.  
 AC AAP50439;  
 DT 01-JAN-1980 (first entry)  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (Ltn)/heat-stable toxin (ST) polypeptide repeating units.



XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX Synthetic.  
 OS  
 XX W08502611-A.  
 XX  
 XX 20-JUN-1985.  
 PD  
 XX  
 XX 12-DEC-1984; 84WO-US02030.  
 PF  
 XX 12-DEC-1983; 83US-0559469.  
 PR  
 XX  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 PA  
 XX Houghten RA;  
 PI  
 XX WPI; 1985-159230/26.  
 DR  
 XX  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 XX Claim 8; Page 100; 120pp; English.  
 PS  
 XX  
 XX The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 CC  
 XX  
 SQ Sequence 41 AA;

Query Match 87.9%; Score 94; DB 6; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOK 18  
 |||||  
 DB 9 GETFOVEVPGSOHIDSOK 26

RESULT 9  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX  
 XX AAP50436;  
 AC  
 XX  
 XX 01-JAN-1980 (first entry)  
 DT  
 XX  
 XX Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 DE  
 XX  
 XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 KW  
 XX  
 XX Synthetic.  
 OS  
 XX W08502611-A.  
 PN  
 XX  
 XX 20-JUN-1985.  
 PD  
 XX  
 XX 12-DEC-1984; 84WO-US02030.  
 PF  
 XX 12-DEC-1983; 83US-0559469.  
 PR  
 XX  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 PA  
 XX Houghten RA;  
 PI  
 XX WPI; 1985-159230/26.  
 DR  
 XX  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT

PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 XX Claim 8; Page 100; 120pp; English.  
 PS  
 XX  
 XX The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 CC  
 XX  
 SQ Sequence 46 AA;

Query Match 82.2%; Score 88; DB 6; Length 46;  
 Best Local Similarity 94.4%; Pred. No. 1.2e-07;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOK 18  
 |||||  
 DB 11 GXTFOVEVPGSOHIDSOK 28

RESULT 10  
 AAP30600  
 ID AAP30600 standard; protein; 47 AA.  
 XX  
 XX AAP30600;  
 AC  
 XX  
 XX 21-APR-1992 (first entry)  
 DT  
 XX  
 XX Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 35, 67 and 73.  
 DE  
 XX  
 XX Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 KW  
 XX  
 XX Vibrio cholerae.  
 OS  
 XX  
 XX EP95426-A.  
 PN  
 XX  
 XX 30-NOV-1983.  
 PD  
 XX  
 XX 26-MAY-1983; 83EP-0401052.  
 PF  
 XX  
 XX 26-MAY-1982; 82FR-0009167.  
 PR  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 PA  
 XX  
 XX Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 PI  
 XX  
 XX WPI; 1983-834645/49.  
 DR  
 XX  
 XX Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 PT  
 XX  
 XX Claim 8; Page 11; 13pp; French.  
 PS  
 XX  
 XX The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against, cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 CC  
 XX  
 SQ Sequence 47 AA;

Query Match 81.8%; Score 87.5; DB 4; Length 47;  
 Best Local Similarity 86.4%; Pred. No. 1.4e-07;  
 Matches 19; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 1 GEFQO-VEVPGSQHIDSOKKAI 21  
 |||:|||||  
 Db 16 GATFEVVEVPGSQHIDSOKKAI 37

RESULT 11  
 AAP30265  
 ID AAP30265 standard; Protein; 26 AA.  
 XX  
 AC AAP30265;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which  
 carries an Arg at posns. 67 and 73.  
 DE  
 XX  
 KM Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82FR-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 DR WPI; 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 7; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and Vaccination against cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 CC  
 SQ Sequence 26 AA;  
 XX

Query Match 74.8%; Score 80; DB 4; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VEVPGSQHIDSOKKAI 21  
 |||||||  
 Db 1 VEVPGSQHIDSOKKAI 16

RESULT 12  
 AAP93498  
 ID AAP93498 standard; protein; 15 AA.  
 XX  
 AC AAP93498;

XX 03-MAY-1990 (first entry)  
 DT  
 XX CTP3 epitope of the Cholera toxin B subunit.  
 DE  
 XX CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;  
 KM vaccine; immunotherapy; ds;  
 XX  
 PN WO8910967-A.  
 PN  
 PD 16-NOV-1989.  
 PD  
 PF 05-MAY-1989; 89WO-US01932.  
 PF  
 PR 05-MAY-1988; 88US-0190570.  
 PR  
 XX (PRAXIS-) PRAXIS BIOLOGICS INC.  
 PA (STRD ) LEYLAND STANDFORD JUNIOR UNIV.  
 XX  
 PI Marjarian WR, Stocker BAD, Newton SMC;  
 DR WPI; 1989-356496/48.  
 DR N-PSDB; AAN92414.  
 XX  
 PT New recombinant flagellin gene including sequence - for heterologous  
 PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.  
 PT of antibodies.  
 XX  
 PS Disclosure; fig.4B; 137pp; English.  
 XX  
 CC This sequence corresponds to the CTP3 epitope of the Cholera toxin B  
 CC subunit. The DNA sequence encoding this ligates to othersynthetic  
 CC oligonucleotides to form a new recombinant gene. This encodes  
 CC a flagellin fusion protein which can be used in vaccines for immuno-  
 CC therapy.  
 CC  
 SQ Sequence 15 AA;  
 XX

Query Match 71.0%; Score 76; DB 10; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VEVPGSQHIDSOKKA 20  
 |||||||  
 Db 1 VEVPGSQHIDSOKKA 15

RESULT 13  
 AAR85125  
 ID AAR85125 standard; peptide; 15 AA.  
 XX  
 AC AAR85125;  
 XX  
 DT 13-JUN-1996 (first entry)  
 DT  
 XX Cholera toxin B antigenic peptide fragment CTP3.  
 DE  
 XX Conjugate; cholera; B toxin; peptide fragment; microparticulate;  
 KM inert carrier; modified silica; thyroglobulin; oral vaccine;  
 KM immunisation; infection; insoluble; digestive tract; antigen;  
 KM intestines; antibodies; secretory; Iga class.  
 XX  
 OS Vibrio cholerae.  
 OS  
 PN WO9529701-A1.  
 PN  
 PD 09-NOV-1995.  
 PD  
 PF 02-MAY-1995; 95WO-EP01661.  
 PF  
 PR 03-MAY-1994; 94IL-0109519.  
 PR  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 PA

```

XX Marks RS, Mirelman D, Sela M;
XX WPI: 1995-403805/51.
XX
XX Vaccines for oral immunisation against infecting agents, e.g.
XX cholera - comprise a conjugate of an antigen of an infecting agent
XX covalently bound to micro:particulate inert carrier, e.g. modified
XX aldehyde silica
XX
XX Claim 7; Page 25; 40pp; English.
XX
XX A compsn. comprising a conjugate of an antigenic cholera B toxin
XX peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently
XX bound to a microparticulate inert carrier (e.g. modified silica or
XX thyroglobulin) can be used as an oral vaccine for immunisation
XX against cholera infection. The inert carrier is insoluble in the
XX digestive tract, allowing presentation of the antigen in the
XX intestines, where it will elicit antibodies mainly of the
XX secretory IGA class.
XX
XX Sequence 15 AA;
XX
XX Query Match 71.0%; Score 76; DB 16; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 VEVPGSOHIDSOKKA 20
Db 1 VEVPGSOHIDSOKKA 15

```

```

RESULT 14
AAR76748
ID AAR76748 standard; Protein; 23 AA.
XX
XX AAR76748;
XX
XX 18-MAR-1996 (first entry)
XX
XX Residues 50-64 of cholera toxin B subunit and FimH 224-226.
XX
XX FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
XX FimH; FimH; receptor binding site; PCR; amplify; ss.
XX
XX Chimeric - Vibrio cholerae.
XX OS Chimeric - Escherichia coli.
XX
XX Key Location/Qualifiers
XX FTH Misc-difference 1
XX FT Peptide /note= "Represents FimH residue 224"
XX FT Peptide 2..4
XX FT Peptide /note= "Linker peptide"
XX FT Peptide 5..19
XX FT Peptide /note= "Cholera toxin B subunit 50-64"
XX FT Peptide 20..22
XX FT Misc-difference 23
XX FT /note= "Linker peptide"
XX FT /note= "Represents FimH residue 226"
XX
XX W09520657-A1.
XX
XX 03-AUG-1995.
XX
XX 27-JAN-1995; 95WO-DK00042.
XX
XX 27-JAN-1994; 94US-0187166.
XX
XX (GABI-) GX BIOSYSTEMS AS.
XX
XX Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX WPI: 1995-275442/36.
XX

```

```

DR N-PSDB; AA093061.
XX
XX Receptor specific bacterial adhesins - useful for targetting active
XX compounds and microbial cells to locations of receptors
XX
XX Example 3; Page 58; 152pp; English.
XX
XX This sequence is encoded by a fragment of the the plasmid pLPA93
XX which was used in the production of fimH fusion genes comprising
XX the cholera toxin B subunit inserted into the fimH gene. This insert
XX shows the inclusion of the B subunit into the FimH protein at position
XX 224-226. The chimeric genes were then opt. further modified by insertion
XX of the hepatitis B virus surface antigen pre-S2 region into a different
XX position of the FimH adhesin of type 1 fimbriae. Restriction site handles
XX (BglII-sites) were introduced into the fimH gene, and the foreign
XX epitopes are then inserted in-frame. In the selected positions the
XX insertion of the epitopes did not significantly alter the adhesive
XX function of the FimH protein. The expression of the chimeric proteins
XX on the surface of fimbriae on bacterial hosts illustrated the possibility
XX of using bacterial adhesins as general presenters of foreign antigens and
XX epitopes. These chimeric genes may be used in the production of variant
XX FimH adhesins which may be useful for targetting active compounds
XX and microbial cells to locations comprising selected receptors to which
XX the adhesins bind.
XX
XX Sequence 23 AA;
XX
XX Query Match 71.0%; Score 76; DB 16; Length 23;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 VEVPGSOHIDSOKKA 20
Db 5 VEVPGSOHIDSOKKA 19

```

```

RESULT 15
AA87461
ID AA87461 standard; peptide; 12 AA.
XX
XX AA87461;
XX
XX 03-JUL-2000 (first entry)
XX
XX Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.
XX
XX Cholera toxin subunit B; CTxB; heat labile enterotoxin subunit B; EtxB;
XX beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX adjuvant; immune disorder; diarrhoea.
XX
XX Vibrio cholerae.
XX OS Escherichia coli.
XX
XX W0200014114-A1.
XX
XX 16-MAR-2000.
XX
XX 07-SEP-1999; 99WO-GB02970.
XX
XX 07-SEP-1998; 98GB-0019484.
XX
XX (UYBR-) UNIV BRISTOL.
XX
XX Williams NA, Hirst TR;
XX WPI: 2000-256943/22.
XX
XX Derivatives of Escherichia coli heat labile enterotoxins useful as
XX immunomodulators and for treating diarrhoea and which do not bind the
XX glycolipid receptor GM-1 -
XX Disclosure; Page 15; 62pp; English.
XX

```

CC The invention relates to peptide fragments of the Escherichia coli heat  
CC labile enterotoxin (Etx) and its closely related homologues, cholera  
CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
CC composed of one A subunit and five identical B subunits. The A subunit  
CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
CC facilitate the entry of subunit A into the host cell via the binding and  
CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
CC for some of the effects of Etx and Ctx, it has been found that certain  
CC effects of the toxins, such as immunomodulation, are not mediated  
CC through GM-1 binding. The peptides of the invention are fragments of the  
CC beta-4' alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
CC Therefore, the peptides may be used in the production of a composition  
CC for treating, preventing and/or modulating a disease associated with an  
CC immune disorder and/or toxin-induced diarrhoea. Sequences AA187460-187463  
CC represent preferred peptides of the invention, AA187460 being  
CC particularly preferred.

XX  
SQ Sequence 12 AA;

Query Match 57.9%; Score 62; DB 21; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.00065;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 VEVPGSQHDSQ 17  
|  
|  
|  
|  
|  
|  
|  
|  
|  
|  
|  
|  
Db 1 VEVPGSQHDSQ 12

Search completed: October 24, 2002, 15:35:04  
Job time : 32.7049 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:33:23 ; Search time 12.0492 Seconds  
(without alignments)  
42.570 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEYPSQHIHSQKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 177902

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	93	2	US-08-292-968-26 Sequence 26, Appl
2	107	100.0	93	2	US-08-467-974-26 Sequence 26, Appl
3	107	100.0	93	2	US-08-467-536-26 Sequence 26, Appl
4	107	100.0	93	3	US-08-467-976-26 Sequence 26, Appl
5	107	100.0	93	4	US-09-082-514-26 Sequence 26, Appl
6	39	36.4	75	1	US-08-350-884-35 Sequence 35, Appl
7	39	36.4	75	1	US-08-709-173-35 Sequence 35, Appl
8	39	36.4	75	2	US-08-833-678A-3 Sequence 3, Appl
9	39	36.4	75	4	US-08-529-169A-3 Sequence 3, Appl
10	39	36.4	76	1	US-08-232-815-2 Sequence 2, Appl
11	39	36.4	76	1	US-08-350-906-2 Sequence 2, Appl
12	39	36.4	76	1	US-08-468-274-2 Sequence 2, Appl
13	39	36.4	76	5	PCT-US95-04536-2 Sequence 2, Appl
14	39	36.4	76	1	US-08-450-834-6 Sequence 6, Appl
15	38	35.5	77	6	5510474-3 Patent No. 5510474
16	38	35.5	78	3	US-08-505-486-94 Sequence 94, Appl
17	38	35.5	78	3	US-08-801-028-94 Sequence 94, Appl
18	38	35.5	78	3	US-09-340-154-94 Sequence 94, Appl
19	38	35.5	78	5	PCT-US95-09338-94 Sequence 94, Appl
20	38	35.5	78	5	PCT-US95-09339-94 Sequence 94, Appl
21	38	35.5	78	5	PCT-US95-09339-94 Sequence 94, Appl
22	36	33.6	27	3	US-09-045-632-83 Sequence 83, Appl
23	36	33.6	28	2	US-08-726-306A-101 Sequence 101, Appl
24	36	33.6	79	3	US-09-100-804-23 Sequence 15, Appl
25	34	31.8	93	4	US-09-069-023-15 Sequence 61, Appl
26	33	30.8	28	3	US-09-045-632-61 Sequence 5, Appl
27	33	30.8	98	3	US-09-045-632-5 Sequence 5, Appl

28	32.5	30.4	21	2	US-08-308-818-8
29	32	29.9	37	4	US-08-817-787-3
30	32	29.9	37	4	US-08-817-787-10
31	32	29.9	41	6	5208144-18
32	32	29.9	76	2	US-09-070-060-8
33	32	29.9	76	3	US-09-357-746-8
34	32	29.9	76	4	US-08-817-787-1
35	32	29.9	76	4	US-08-817-787-8
36	32	29.9	76	4	US-09-100-802-1
37	31.5	29.4	35	1	US-08-418-893D-16
38	31	29.0	21	1	US-08-355-888A-14
39	31	29.0	21	3	US-08-693-697-14
40	31	29.0	21	2	US-08-693-696-14
41	31	29.0	28	1	US-08-255-558B-6
42	31	29.0	66	2	US-08-849-480A-10
43	31	29.0	75	2	US-08-332-562A-29
44	31	29.0	76	2	US-08-284-391B-53
45	31	29.0	76	4	US-09-218-950-53

#### ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 5856122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDNESS: single  
TOPOLOGY: linear

Sequence 8, Appl  
Sequence 3, Appl  
Sequence 10, Appl  
Patent No. 5208144  
Sequence 8, Appl  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 16, Appl  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 10, Appl  
Sequence 29, Appl  
Sequence 53, Appl

US-08-292-968-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 35 GETFOVEVPGSQHIDSOKKAI 55

## RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,974  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/467,536  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-454 MIS:v9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEO ID NO: 26;  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-467-974-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GETFOVEVPGSQHIDSOKKAI 21

Db 35 GETFOVEVPGSQHIDSOKKAI 55

## RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,536  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-455 MIS:v9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEO ID NO: 26;  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-467-536-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 35 GETFOVEVPGSQHIDSOKKAI 55

RESULT 4  
US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022  
; GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,976  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-453 MIS-VG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-976-26  
Query Match 100.0%; Score 107; DB 3; Length 93;  
Best local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 35 GETFOVEVPGSHIDSOKKAI 21  
DB 35 GETFOVEVPGSHIDSOKKAI 55  
RESULT 5  
US-09-082-514-26  
Sequence 26, Application US/09082514  
Patent No. 6168928  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,514  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 24-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-810  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-082-514-26  
Query Match 100.0%; Score 107; DB 4; Length 93;  
Best local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 35 GETFOVEVPGSHIDSOKKAI 21  
DB 35 GETFOVEVPGSHIDSOKKAI 55  
RESULT 6  
US-08-350-884-35  
Sequence 35, Application US/08350884  
Patent No. 5585258  
GENERAL INFORMATION:  
APPLICANT: HOGGHTON, MICHAEL  
APPLICANT: CHOO, QUI LIM  
APPLICANT: KOO, GEORGE  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FORSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,884  
FILING DATE: 06-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/680,296  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:

Query Match	36.4%;	Score 39;	DB 1;	Length 75
Best Local Similarity	42.9%;	Pred. No. 11;		
Matches	9;	Conservative	3;	Mismatches 9; Indels

1  
 2 TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
 3  
 4 TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY  
 5  
 6 NUMBER OF SEQUENCES: 6  
 7  
 8 CORRESPONDENCE ADDRESS:  
 9  
 10 ADDRESSEE: CHIRON CORPORATION



US-08-529-169A-3  
Sequence 3, Application US/08529169A  
Patent No. 6194140  
GENERAL INFORMATION:  
APPLICANT: HOOCH, QUI-LIM  
APPLICANT: CHOO, QUI-LIM  
APPLICANT: HAN, JANG  
APPLICANT: CHOE, JOONHO  
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING  
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: Intellectual Property - R440, P.O. Box 8097  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,169A  
FILING DATE: 15-SEP-1995  
CLASSIFICATION: 4325  
ATTORNEY/AGENT INFORMATION:  
NAME: Hardin, Alisa A.

```

? Sequence 2 Application US/08232815
? Patent No. 5503977
?
? GENERAL INFORMATION:
?
? APPLICANT: Johnson, Nils
? APPLICANT: Varshavsky, Alexander
? TITLE OF INVENTION: SPLIT UBIQUITIN PROTEIN SENSOR
? NUMBER OF SEQUENCES: 6
?
? CORRESPONDENCE ADDRESS:
?
? ADDRESSEE: Kevin M. Farrell
? STREET: P.O. Box 999
? CITY: York Harbor
? STATE: Maine
? COUNTRY: USA
?
? ZIP: 03911
?
? COMPUTER READABLE FORM:
?

```

```

      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US//08/232,815
      FILING DATE:  22-APR-1994
      CLASSIFICATION:  435
      ATTORNEY/AGENT INFORMATION:
      NAME:  Farrell, Kevin M.
      REGISTRATION NUMBER:  35,505
      REFERENCE/DOCKET NUMBER:  CIT-9303
      TELECOMMUNICATION INFORMATION:
      TELEPHONE:  207 363-0558
      TELEFAX:  207 363-0528
      INFORMATION FOR SEQ ID NO:  2:
      SEQUENCE CHARACTERISTICS:
      LENGTH:  76 amino acids
      TYPE:  amino acid
      TOPOLOGY:  linear
      MOLECULE TYPE:  protein
      US-08-232-815-2

Query Match      36.4%;   Score 39;   DB 1;   Length 76;
Best Local Similarity  42.9%;   Pred. NO. 11;
Matches      9;   Conservative      3;   Mismatches      9;   Indels      0;   Gaps      0;

```



QY 1 GETFOVEVPGSQHDSQKAI 21  
 DB 10 GKTITLEVSSDPTIDNVKSKI 30

## RESULT 15

US-08-450-834-6  
 ; Sequence 6, Application US/08450834  
 ; Patent No. 5773705  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vierstra, Richard D  
 ; APPLICANT: Hondred, David  
 ; APPLICANT: Callis, Judy  
 ; TITLE OF INVENTION: Ubiquitin Fusion Protein System for  
 ; TITLE OF INVENTION: Protein Production in Plants  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: P.O. Box 2113  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: USA  
 ; ZIP: 53701-2113  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/450,834  
 ; FILING DATE: 25-MAY-1995  
 ; CLASSIFICATION: 800  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/999,709  
 ; FILING DATE: 31-DEC-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J  
 ; REGISTRATION NUMBER: 27,386  
 ; REFERENCE/DOCKET NUMBER: 960296.92425  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 608-251-5000  
 ; TELEFAX: 608-251-9166  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 76 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-450-834-6

Query Match 35.5%; Score 38; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 17;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 GETFOVEVPGSQHDSQKAI 21  
 DB 10 GKTITLEVSSDPTIDNVKSKI 30

Search completed: October 24, 2002, 15:38:24  
 Job time : 12.0492 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:09 : Search time 139.77 Seconds  
(without alignments)  
52.884 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107  
Sequence: 1 GETFOVEVPSQHIDSKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 2671256

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/PCUTUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US10\_COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US10\_COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	21	US-09-786-648-5	Sequence 5, Appli
2	107	100.0	93	US-08-110-947-10	Sequence 10, Appli
3	107	100.0	93	US-08-110-947A-26	Sequence 26, Appli
4	107	100.0	93	US-08-251-121-26	Sequence 26, Appli
5	101	94.4	21	US-09-786-648-4	Sequence 4, Appli
6	76	71.0	15	US-08-732-371-1	Sequence 1, Appli
7	76	71.0	15	US-08-732-371A-1	Sequence 1, Appli

8	62	57.9	12	US-09-786-648-3	Sequence 3, Appli
9	46	43.0	21	US-09-708-427-6862	Sequence 6862, Ap
10	45	42.1	91	US-09-758-453-761	Sequence 761, App
11	43	40.2	61	US-09-708-427-44143	Sequence 44143, A
12	43	40.2	76	US-09-708-427-44142	Sequence 44142, A
13	43	40.2	84	US-60-188-162-4087	Sequence 4087, A
14	41	38.3	51	US-09-866-066-35	Sequence 35, Appli
15	41	38.3	91	US-09-595-298A-470	Sequence 470, App
16	40	37.4	92	US-09-800-971-11	Sequence 11, Appli
17	39	36.4	7	US-09-786-648-2	Sequence 2, Appli
18	39	36.4	14	US-09-072-955A-27	Sequence 27, Appli
19	39	36.4	73	US-07-680-296-35	Sequence 35, Appli
20	39	36.4	75	US-08-833-678-3	Sequence 3, Appli
21	39	36.4	75	US-09-483-799-3	Sequence 3, Appli
22	39	36.4	75	US-09-884-455-35	Sequence 35, Appli
23	39	36.4	75	US-09-884-456-35	Sequence 35, Appli
24	39	36.4	85	US-60-177-646-3942	Sequence 3942, Ap
25	38	35.5	35	US-09-260-687-7	Sequence 7, Appli
26	38	35.5	35	US-09-516-277-7	Sequence 11, Appli
27	38	35.5	35	US-09-516-277-11	Sequence 11, Appli
28	38	35.5	35	US-09-516-277-11	Sequence 11, Appli
29	38	35.5	50	US-09-620-111B-3048	Sequence 3048, Ap
30	38	35.5	50	US-09-620-111B-3049	Sequence 3049, Ap
31	38	35.5	51	PCT-US01-00663-37905	Sequence 37905, A
32	38	35.5	51	US-09-864-761-41979	Sequence 41979, A
33	38	35.5	65	US-09-260-687-10	Sequence 10, Appli
34	38	35.5	65	US-09-516-277-10	Sequence 10, Appli
35	38	35.5	69	US-09-708-427-37221	Sequence 37221, A
36	38	35.5	71	US-09-708-427-75825	Sequence 75825, A
37	38	35.5	73	US-09-248-796-18744	Sequence 18744, A
38	38	35.5	73	US-60-096-409-18744	Sequence 18744, A
39	38	35.5	76	US-07-999-709-6	Sequence 6, Appli
40	38	35.5	76	US-09-260-687-6	Sequence 6, Appli
41	38	35.5	76	US-09-516-277-6	Sequence 6, Appli
42	38	35.5	76	US-09-620-111B-3047	Sequence 3047, Ap
43	38	35.5	78	US-09-857-841-2	Sequence 2, Appli
44	38	35.5	78	US-09-482-611-94	Sequence 94, Appli
45	38	35.5	78	US-09-482-611B-94	Sequence 94, Appli

#### ALIGNMENTS

RESULT 1  
US-09-786-648-5  
; Sequence 5, Application US/09786648  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Neil Andrew  
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci  
; FILE REFERENCE: 7438  
; CURRENT APPLICATION NUMBER: US/09/786,648  
; PRIORITY FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: PCT/GB99/02970  
; PRIOR FILING DATE: 1999-09-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: MS DOS  
; SEQ ID NO 5  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: E. coli  
; FEATURE:  
; LOCATION: 45...65  
; OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivabl  
US-09-786-648-5

Query Match 100.0%; Score 107; DB 21; Length 21;  
Best local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GETFOVEVPSQHIDSOKKAI 21  
|||||  
RESULT 2  
US-08-110-947-10  
; Sequence 10, Application US/08110947  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: OOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson  
; STREET: Davis Hwy.  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/110,947  
; FILING DATE: 24-AUG-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FALLOW, Charles W  
; REGISTRATION NUMBER: 28,946  
; REFERENCE/DOCKET NUMBER: 1038-303 MTS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0810  
; TELEX: 89-9456 LUKPAT  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-110-947-10  
Query Match 100.0%; Score 107; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 9.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPSQHIDSOKKAI 21  
|||||  
Db 35 GETFOVEVPSQHIDSOKKAI 55  
|||||  
RESULT 3  
US-08-110-947A-26  
; Sequence 26, Application US/08110947A  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: OOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson  
; STREET: Davis Highway  
; CITY: Arlington

STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,947A  
FILING DATE: 24-AUG-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLOW, Charles W  
REGISTRATION NUMBER: 28,946  
REFERENCE/DOCKET NUMBER: 1038-303 MTS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0810  
TELEX: 89-9456 LUKPAT  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-110-947A-26  
Query Match 100.0%; Score 107; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 9.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPSQHIDSOKKAI 21  
|||||  
Db 35 GETFOVEVPSQHIDSOKKAI 55  
|||||  
RESULT 4  
US-08-251-121-26  
; Sequence 26, Application US/08251121  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: OOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; APPLICANT: ARMSTRONG, Glen D  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/251,121  
; FILING DATE: 31-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-335

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-251-121-26

Query Match 100.0%; Score 107; DB 6; Length 93;  
Best Local Similarity 100.0%; Pred. No. 9.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSQKAI 21  
|||||

Db 35 GETFOVEVPGSOHIDSQKAI 55

RESULT 5  
US-09-786-648-4  
Sequence 4, Application US/097866648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine  
TITLE OF INVENTION: Adjuvants  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/GB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 4  
LENGTH: 21  
TYPE: PPT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 45...65  
OTHER INFORMATION: Isolated or synthetic Etxb beta4-alpha2 loop fragment derivable from  
US-09-786-648-4

Query Match 94.4%; Score 101; DB 21; Length 21;  
Best Local Similarity 95.2%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSQKAI 21  
|||||

Db 1 GATFOVEVPGSOHIDSQKAI 21

RESULT 6  
US-08-732-371-1  
Sequence 1, Application US/08732371  
GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
APPLICANT: SELA, Michael  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
TITLE OF INVENTION: INFECTING AGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371-1

Query Match 71.0%; Score 76; DB 11; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.6e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VEVPQSOHIDSQKKA 20  
|||||

Db 1 VEVPQSOHIDSQKKA 15

RESULT 7  
US-08-732-371A-1  
Sequence 1, Application US/08732371A  
GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
APPLICANT: SELA, Michael  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
TITLE OF INVENTION: INFECTING AGENTS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371A  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371A-1

Query Match  
Best Local Similarity 71.0%; Score 76; DB 11; Length 15;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VEPGSGHDSOKKA 20  
Db 1 VEPGSGHDSOKKA 15

RESULT 8  
US-09-786-648-3  
Sequence 3, Application US/09786648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
APPLICANT: Hirst, Timothy Raymond  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/GB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 50...61  
OTHER INFORMATION: Isolated or synthetic EtXb beta4-alpha2 loop fragment derivable  
US-09-786-648-3

Query Match  
Best Local Similarity 57.9%; Score 62; DB 21; Length 12;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VEPGSGHDSO 17  
Db 1 VEPGSGHDSO 12

RESULT 9  
US-09-708-427-6862  
Sequence 6862, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 6862  
LENGTH: 91  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..91  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1..91  
OTHER INFORMATION: Ceres Seq. ID 1816431  
US-09-708-427-6862

Query Match  
Best Local Similarity 43.0%; Score 46; DB 21; Length 91;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHDSOKKAI 21  
Db 58 GTFVMEVKGSEIIQOVKNMI 78

RESULT 10  
US-09-758-463-761  
Sequence 761, Application US/09758463  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM01  
CURRENT APPLICATION NUMBER: US/09/758,463  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 761  
LENGTH: 99  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (9)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-463-761

Query Match  
Best Local Similarity 42.1%; Score 45; DB 21; Length 99;  
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHDSOKKAI 21  
Db 74 GTFVVGAGSGSEGESEKAI 94

RESULT 11  
US-09-708-427-44143  
Sequence 44143, Application US/09708427  
GENERAL INFORMATION:  
APPLICANT: N. ALEXANDROV et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE REFERENCE: 2750-1243P  
CURRENT APPLICATION NUMBER: US/09/708,427  
CURRENT FILING DATE: 2000-11-09  
NUMBER OF SEQ ID NOS: 85364  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 44143  
LENGTH: 61  
TYPE: PRT  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..61  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1..61  
OTHER INFORMATION: Ceres Seq. ID 1848967  
US-09-708-427-44143

Query Match  
Best Local Similarity 40.2%; Score 43; DB 21; Length 61;  
Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 1 GETFOVEVPGSHDSOKK 19



Db 30 GDTF----PGSEHLPADRK 44

RESULT 12

US-09-708-427-44142

; Sequence 44142, Application US/09708427

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; FILE REFERENCE: 2750-1243P

; CURRENT APPLICATION NUMBER: US/09/708,427

; CURRENT FILING DATE: 2000-11-09

; NUMBER OF SEQ ID NOS: 85364

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 44142

; LENGTH: 76

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1..76

; OTHER INFORMATION: Xaa is any amino acid

; NAME/KEY: misc\_feature

; LOCATION: 1..76

; OTHER INFORMATION: Ceres Seq. ID 1848966

US-09-708-427-44142

Query Match

Best Local Similarity 40.2%; Score 43; DB 21; Length 76;

Best Local Similarity 42.1%; Pred. No. 48;

Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 1 GETFOVEVPGSOHDSOKR 19

Db 45 GDTF----PGSEHLPADRK 59

RESULT 13

US-60-188-162-4087

; Sequence 4087, Application US/60188162

; GENERAL INFORMATION:

; APPLICANT: Beasley, Ellen

; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

; FILE REFERENCE: CLO00518

; CURRENT APPLICATION NUMBER: US/60/188,162

; CURRENT FILING DATE: 2000-05-09

; NUMBER OF SEQ ID NOS: 5094

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4087

; LENGTH: 84

; TYPE: PRT

; ORGANISM: HUMAN

US-60-188-162-4087

Query Match

Best Local Similarity 40.2%; Score 43; DB 26; Length 84;

Best Local Similarity 53.3%; Pred. No. 55;

Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHID 15

Db 42 GSTSTREYVPGOSHHD 56

RESULT 14

US-09-866-066-35

; Sequence 35, Application US/09866066

; GENERAL INFORMATION:

; APPLICANT: Benjamin, Christopher

; APPLICANT: Roberds, Steve

; APPLICANT: Rubie, Cara

US-09-866-066-35

Query Match

Best Local Similarity 38.3%; Score 41; DB 19; Length 91;

Best Local Similarity 35.3%; Pred. No. 1.3e+02;

Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 EFTFOVEVPGSOHDSOKR 18

Db 37 QSLIVKVPGRKQVSEKE 53

; APPLICANT: Gotow, Lisa

; APPLICANT: Karnovsky, Alla

; TITLE OF INVENTION: Human Ion Channels

; FILE REFERENCE: 00229.051

; CURRENT APPLICATION NUMBER: US/09/866,066

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: 60/207,152

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/207,257

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/207,119

; PRIOR FILING DATE: 2000-05-26

; NUMBER OF SEQ ID NOS: 42

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 35

; LENGTH: 51

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-866-066-35

Query Match

Best Local Similarity 38.3%; Score 41; DB 22; Length 51;

Best Local Similarity 58.3%; Pred. No. 63;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 FOVEVPGSOHID 15

Db 5 FPEIIPGSSHWD 16

RESULT 15

US-09-595-298A-470

; Sequence 470, Application US/09595298A

; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nickolai

; APPLICANT: BROVER, Vyacheslav

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept

; FILE REFERENCE: 2750-0953P

; CURRENT APPLICATION NUMBER: US/09/595,298A

; CURRENT FILING DATE: 2000-06-16

; NUMBER OF SEQ ID NOS: 2756

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 470

; LENGTH: 91

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: peptide

; LOCATION: (1)..(91)

; OTHER INFORMATION: Ceres Seq. ID no. 1024399

; NAME/KEY: misc\_feature

; LOCATION: (1)..(1)

; OTHER INFORMATION: Xaa is any aa, unknown or other

US-09-595-298A-470

Query Match

Best Local Similarity 38.3%; Score 41; DB 19; Length 91;

Best Local Similarity 35.3%; Pred. No. 1.3e+02;

Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 EFTFOVEVPGSOHDSOKR 18

Db 37 QSLIVKVPGRKQVSEKE 53

Search completed: October 24, 2002, 15:45:18  
Job time : 140.77 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:39 : Search time 45.7669 Seconds  
(without alignments)  
137.008 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEVPGSQHIDSOKKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 378271

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/US01\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US01\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	42.1	84	US-09-620-393B-3233	Sequence 3233, Ap
2	45	42.1	99	US-10-217-607-761	Sequence 761, App
3	43	40.2	76	US-09-791-537-81337	Sequence 81337, A
4	40	37.4	76	US-09-791-537-137751	Sequence 137751, A
5	39	36.4	45	US-09-791-537-151859	Sequence 151859, A
6	39	36.4	60	US-10-219-999-51216	Sequence 51216, A
7	39	36.4	62	US-09-791-537-120702	Sequence 120702, A
8	39	36.4	62	US-09-791-537-120704	Sequence 120704, A
9	39	36.4	75	US-10-232-643-3	Sequence 3, Appli
10	39	36.4	76	US-09-791-537-81370	Sequence 81370, A
11	39	36.4	77	US-09-791-537-140209	Sequence 140209, A
12	39	36.4	87	US-10-155-881-12211	Sequence 12211, A
13	38	35.5	41	US-10-219-999-59647	Sequence 59647, A
14	38	35.5	48	US-09-791-537-74565	Sequence 74565, A
15	38	35.5	51	US-10-182-995-28451	Sequence 28451, A
16	38	35.5	51	US-10-182-995-28451	Sequence 28451, A
17	38	35.5	51	US-10-203-136-37757	Sequence 37757, A
18	38	35.5	51	US-10-203-136-37757	Sequence 37757, A
19	38	35.5	51	US-10-182-997-26224	Sequence 26224, A
20	38	35.5	51	US-10-203-135-36160	Sequence 36160, A
21	38	35.5	51	US-10-203-137-37905	Sequence 37905, A
22	38	35.5	51	US-10-203-139-36373	Sequence 36373, A
23	38	35.5	56	US-09-791-537-74562	Sequence 74562, A
24	38	35.5	69	US-09-513-996A-39502	Sequence 39502, A
25	38	35.5	69	US-09-513-996A-50521	Sequence 50521, A
26	38	35.5	70	US-09-620-393B-4316	Sequence 4316, Ap

## ALIGNMENTS

27	38	35.5	71	5	US-09-513-996A-40667	Sequence 40667, A
28	38	35.5	72	5	US-09-791-537-110433	Sequence 110433, A
29	38	35.5	73	5	US-09-620-393B-5363	Sequence 5363, Ap
30	38	35.5	76	5	US-09-791-537-81335	Sequence 81335, A
31	38	35.5	76	5	US-09-791-537-81362	Sequence 81362, A
32	38	35.5	76	6	US-10-219-999-34217	Sequence 34217, A
33	38	35.5	77	5	US-09-620-393B-870	Sequence 870, App
34	38	35.5	77	5	US-09-620-393B-5214	Sequence 5214, Ap
35	38	35.5	77	5	US-09-935-625-2479	Sequence 2479, Ap
36	38	35.5	77	5	US-09-935-625-6721	Sequence 6721, Ap
37	38	35.5	77	5	US-09-935-625-11032	Sequence 11032, A
38	38	35.5	77	5	US-09-935-625-15721	Sequence 15721, A
39	38	35.5	77	5	US-09-935-625-18647	Sequence 18647, A
40	38	35.5	77	5	US-09-935-625-18895	Sequence 18895, A
41	38	35.5	77	5	US-09-935-625-21350	Sequence 21350, A
42	38	35.5	77	5	US-09-935-625-24220	Sequence 24220, A
43	38	35.5	77	5	US-09-935-625-27445	Sequence 27445, A
44	38	35.5	77	5	US-09-791-537-103174	Sequence 103174, A
45	38	35.5	77	5	US-09-791-537-103413	Sequence 103413, A

RESULT 1  
US-09-620-393B-3233  
Sequence 3233, Application US/09620393B  
GENERAL INFORMATION:  
APPLICANT: ALEXANDROV, Nikolai et al.  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
FILE OF INVENTION: THEREBY  
CURRENT APPLICATION NUMBER: US/09/620,393B  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 9948  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3233  
LENGTH: 84  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc:feature  
LOCATION: 1..84  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc:feature  
LOCATION: 1..84  
OTHER INFORMATION: Ceres Seq. ID 1385062  
US-09-620-393B-3233  
Query Match 42.1%, Score 45; DB 5; Length 84;  
Best Local Similarity 38.1%; Pred. No. 8.1;  
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSQHIDSOKKAI 21  
I:::|||||I:::  
DB 21 GKTSMGPEIDCDENKGL 41  
US-10-217-607-761  
Sequence 761, Application US/10217607  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM0111CN  
CURRENT APPLICATION NUMBER: US/10/217,607  
CURRENT FILING DATE: 2002-08-14  
PRIOR APPLICATION NUMBER: 09/758,463  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628

```

; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 761
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-217-607-761

Query Match
Best Local Similarity 42.1%; Score 45; DB 6; Length 99;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 74 GTTFQVGAGSGESGDSGKAI 94

RESULT 3
US-09-791-537-81337
; Sequence 81337, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 81337
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Dictyostelium discoideum
US-09-791-537-81337

Query Match
Best Local Similarity 40.2%; Score 43; DB 5; Length 76;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 10 GKTITLEVSSDNIENYKAKI 30

RESULT 4
US-09-791-537-137751
; Sequence 137751, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 137751
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-09-791-537-137751

Query Match
Best Local Similarity 37.4%; Score 40; DB 5; Length 76;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 761
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-217-607-761

Query Match
Best Local Similarity 42.1%; Score 45; DB 6; Length 99;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 74 GTTFQVGAGSGESGDSGKAI 94

RESULT 3
US-09-791-537-81337
; Sequence 81337, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 81337
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Dictyostelium discoideum
US-09-791-537-81337

Query Match
Best Local Similarity 40.2%; Score 43; DB 5; Length 76;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 10 GKTITLEVSSDNIENYKAKI 30

RESULT 4
US-09-791-537-137751
; Sequence 137751, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 137751
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-09-791-537-137751

Query Match
Best Local Similarity 37.4%; Score 40; DB 5; Length 76;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 10 GKTITLEVSSDITDNVKKI 30

RESULT 5
US-09-791-537-151859
; Sequence 151859, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 151859
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Phreatamoeba balamuthi
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)..(39)
; OTHER INFORMATION: X is an unknown amino acid
US-09-791-537-151859

Query Match
Best Local Similarity 36.4%; Score 39; DB 5; Length 45;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GTFQVEVPGSQHIDSOKKAI 21
Db 3 GKTITLEVPSDITDNVKKI 23

RESULT 6
US-10-219-999-51216
; Sequence 51216, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 51216
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Zea mays
US-10-219-999-51216

Query Match
Best Local Similarity 47.4%; Score 39; DB 6; Length 60;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 TQVEVPGSQHIDSOKKAI 21
Db 5 TLSVQVFGNHLVFEKKAI 23
```

```

RESULT 7
US-09-791-537-120702
; Sequence 120702, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 120702
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Phreatamoeba balamuthi
US-09-791-537-120702

Query Match
Best Local Similarity 36.4%; Score 39; DB 5; Length 62;
Best Local Similarity 42.9%; Pred. No. 58;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFEVPEPGSOHIDSOKKAI 21
Db 3 GKTITLEVPSDSIDNVKOKI 23

RESULT 8
US-09-791-537-120704
; Sequence 120704, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 120704
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Phreatamoeba balamuthi
US-09-791-537-120704

Query Match
Best Local Similarity 36.4%; Score 39; DB 5; Length 62;
Best Local Similarity 42.9%; Pred. No. 58;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 GETFEVPEPGSOHIDSOKKAI 21
Db 3 GKTITLEVPSDSIDNVKOKI 23

RESULT 9
US-10-232-643-3
; Sequence 3, Application US/10232643
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, OUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097

```

```

; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/232,643
; FILING DATE: 30-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/483,799
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/529,169
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Ailsa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: 0/4
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-232-643-3

Query Match
Best Local Similarity 36.4%; Score 39; DB 6; Length 75;
Best Local Similarity 42.9%; Pred. No. 73;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFEVPEPGSOHIDSOKKAI 21
Db 9 GKTITLEVSSDPTIDNVKSKI 29

RESULT 10
US-09-791-537-81370
; Sequence 81370, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 81370
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Fillobasidiella neoformans var
US-09-791-537-81370

Query Match
Best Local Similarity 36.4%; Score 39; DB 5; Length 76;
Best Local Similarity 42.9%; Pred. No. 74;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFEVPEPGSOHIDSOKKAI 21
Db 10 GKTITLEVSSDPTIDNVKSKI 30

```

RESULT 11  
US-09-791-537-140209  
; Sequence 140209, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomolix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 140209  
; LENGTH: 77  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
US-09-791-537-140209

Query Match 36.4%; Score 39; DB 5; Length 77;  
Best Local Similarity 33.3%; Pred. No. 75;  
Matches 10; Conservative 4; Mismatches 4; Indels 12; Gaps 1;

QY 2 ETEVEVPGS-----QHDSQK 19  
DB 48 EKEVEIPDSAEKLIKEDVYVYIDNLR 77

RESULT 12  
US-10-155-881-12211  
; Sequence 12211, Application US/10155881  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lutfiyya, Linda L.  
; APPLICANT: McIntich, James  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(15300)J  
; CURRENT APPLICATION NUMBER: US/10/155,881  
; CURRENT FILING DATE: 2002-05-22  
; NUMBER OF SEQ ID NOS: 37595  
; SEQ ID NO 12211  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-155-881-12211

Query Match 36.4%; Score 39; DB 6; Length 87;  
Best Local Similarity 38.5%; Pred. No. 87;  
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 5 QVEVPGSQHDSQ 17  
DB 41 ELEYPGAHEDE 53

RESULT 13  
US-10-219-999-59647  
; Sequence 59647, Application US/10219999  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999

; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 59647  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-219-999-59647

Query Match 35.5%; Score 38; DB 6; Length 41;  
Best Local Similarity 42.9%; Pred. No. 52;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETVEVPGSQHDSQKAI 21  
DB 15 GKTILEVSSDTIDNVKAKI 35

RESULT 14  
US-09-791-537-74565  
; Sequence 74565, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomolix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 74565  
; LENGTH: 48  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-791-537-74565

Query Match 35.5%; Score 38; DB 5; Length 48;  
Best Local Similarity 42.9%; Pred. No. 63;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETVEVPGSQHDSQKAI 21  
DB 10 GKTILEVSSDTIDNVKAKI 30

RESULT 15  
US-10-182-995-28451  
; Sequence 28451, Application US/10182995  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: PB 0004 WO 1  
; CURRENT APPLICATION NUMBER: US/10/182,995  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
; NUMBER OF SEQ ID NOS: 29119  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 28451  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL109658.5  
; FEATURE:  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
; FEATURE:  
; OTHER INFORMATION: EST\_HUMAN HIT: AV651727.1, EVALUATE 4.00e-15  
US-10-182-995-28451

Query Match 35.5%; Score 38; DB 6; Length 51;  
Best Local Similarity 46.2%; Pred. No. 68;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 EVPGSQHIDSQK 19  
::|||::|||  
Db 23 DLPGRSHLSKKK 35

Search completed: October 24, 2002, 15:47:34  
Job time : 45.7869 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:31:03 ; Search time 15.8361 Seconds  
(without alignments)  
127.423 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETPQVEVPGSQHIDSOKKAI 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 37652

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	46	43.0	91	2	C96580 hypothetical prote
2	40	37.4	79	2	C95152 IS66 family elemen
3	40	37.4	79	2	H95167 IS66 family elemen
4	39	36.4	77	2	G81388 acyl carrier prote
5	38	35.5	76	1	UQOA ubiquitin - oat
6	38	35.5	76	2	S12114 polyubiquitin - in
7	38	35.5	77	2	S19799 ubiquitin - potato
8	38	35.5	77	2	S16263 ubiquitin precursor
9	38	35.5	79	2	B96034 degenerate transpo
10	38	35.5	80	2	F87520 ubiquitin precursor
11	38	35.5	88	1	UQSV polyubiquitin - ja
12	38	35.5	99	2	S12161 ubiquitin/ribosoma
13	38	35.5	68	2	AG3217 hypothetical prote
14	36	33.6	77	2	S77465 acyl carrier prote
15	36	33.6	91	2	F81109 conserved hypothet
16	35	32.7	41	2	T36863 conserved two-compo
17	35	32.7	78	2	C87457 acyl carrier prote
18	35	32.7	78	2	G69434 conserved hypothet
19	34	31.8	70	2	A12538 hypothetical prote
20	34	31.8	72	2	H69608 carbon storage reg
21	34	31.8	77	1	UQUT ubiquitin precursor
22	34	31.8	78	1	UQUTC ubiquitin precursor
23	34	31.8	99	2	OSBPB7 host specificity p
24	34	31.8	99	2	G84272 50S ribosomal prot
25	33	30.8	65	2	T24159 hypothetical prote
26	33	30.8	70	2	T39058 hypothetical prote
27	33	30.8	77	2	S29238 ubiquitin - Entamo
28	33	30.8	80	2	C41609 acyl carrier prote
29	33	30.8	80	2	C41609

30	33	30.8	82	2	D84387 hypothetical prote
31	33	30.8	95	2	T14743 hypothetical prote
32	32	29.9	35	2	T02897 sigma-E factor reg
33	32	29.9	38	2	F60657 circumsporozoite p
34	32	29.9	51	2	AG2245 hypothetical prote
35	32	29.9	74	1	UQFPM ubiquitin - Medite
36	32	29.9	75	2	A84102 carbon storage reg
37	32	29.9	76	1	UQHU ubiquitin - human
38	32	29.9	76	1	UQBO ubiquitin - bovine
39	32	29.9	76	1	UQUTSF ubiquitin - fall a
40	32	29.9	76	2	A26087 ubiquitin - fruit
41	32	29.9	76	2	S28203 ubiquitin - rabbit
42	32	29.9	76	2	S42750 polyubiquitin UB2
43	32	29.9	77	2	S34334 ubiquitin - Leishm
44	32	29.9	78	2	H71922 acyl carrier prote
45	32	29.9	79	2	S77831 hypothetical prote

#### ALIGNMENTS

RESULT 1  
C96580  
hypothetical protein F1511.6 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96580  
R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaytin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marria  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MIMID:21016719  
A:Accession: C96580  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <SP0>  
A:Cross-references: GB:AE005173; NID:g4587538; PIDN:AAD25769.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F1511.6  
A:Map position: 1  
Query Match 43.0%; Score 46; DB 2; Length 91;  
Best Local Similarity 47.6%; Pred. No. 3.2;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
QY 1 GETPQVEVPGSQHIDSOKKAI 21  
DB 58 GKTFNLEVKSEITIQVKMMI 78  
RESULT 2  
C95152  
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: C95152  
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappel  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MIMID:21357209; PMID:11463916  
A:Accession: C95152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75412.1; PID:g14972794; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI314

Query Match 37.4%; Score 40; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 26;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEPGSOHIDSOKK 19  
Db 26 VEIPGSRPDSKRR 39

## RESULT 3

H95167  
IS66 family element, Orf1 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: H95167  
R:Retelien, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95167

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75537.1; PID:g14972931; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:  
A:Gene: SPI443

Query Match 37.4%; Score 40; DB 2; Length 79;  
Best Local Similarity 50.0%; Pred. No. 26;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 VEPGSOHIDSOKK 19  
Db 26 VEIPGSRPDSKRR 39

## RESULT 4

G81388  
acyl carrier protein Cj0441 [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Nov-2000

C:Accession: G81388

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanyliet, A.; Whitehead, S.; Barré  
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals h  
A:Reference number: A81250; MUID:20150912

A:Accession: G81388

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <PAR>

A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CA874277.1; PID:g696791

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: acpP; Cj0441

C:Superfamily: acyl carrier protein; acyl carrier protein homology

C:Keywords: carrier protein

F:1-72/Domain: acyl carrier protein homology <ACP>

Query Match 36.4%; Score 39; DB 2; Length 77;  
Best Local Similarity 33.3%; Pred. No. 36;  
Matches 10; Conservative 4; Mismatches 4; Indels 12; Gaps 1;

QY 2 ETFOVEVPGS-----OHIDSOKK 19

Db 48 EKFEVEIPDSAEKLIKIEDVWYTDNLKK 77

## RESULT 5

UOQA  
ubiquitin - oat

C:Species: Avena sativa (oat)  
C:Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 16-Aug-1996

C:Accession: A02576

R:Vierstra, R.D.; Langan, S.M.; Schaller, G.E.  
Biochemistry 25, 3105-3108, 1986

A:Title: Complete amino acid sequence of ubiquitin from the higher plant Avena sativa

A:Reference number: A02576

A:Accession: A02576

A:Molecule type: protein

A:Residues: 1-76 <VIE>

C:Superfamily: ubiquitin; ubiquitin homology

C:Keywords: protein degradation

F:1-76/Domain: ubiquitin homology <UBH>

F:76/Cross-link: Isoleptide carboxyl end (gly) (interchain to Lys N6-amino of other p

Query Match 35.5%; Score 38; DB 1; Length 76;  
Best Local Similarity 42.9%; Pred. No. 52;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21  
Db 10 GKTITLEVESSDTIDNVKAKI 30

## RESULT 6

S12114  
polyubiquitin - inky cap (Coprinus congregatus) (fragment)

C:Species: Coprinus congregatus

C:Date: 21-Nov-1993 #sequence\_revision 17-Jul-1998 #text\_change 23-Jul-1999

C:Accession: S12114

R:Foster, L.M.; Loftus, M.G.; Ross, I.K.  
Nucleic Acids Res. 18, 6449, 1990

A:Title: A novel form of ubiquitin found in the basidiomycete fungus, Coprinus congre

A:Reference number: S12114; MUID:91057153

A:Accession: S12114

A:Molecule type: DNA

A:Residues: 1-76 <FOS>

A:Cross-references: EMBL:X54672; NID:g25550; PIDN:CAA38483.1; PID:g2551

A:Experimental source: strain Californian, CC 16 homokaryon, mating type 1

C:Function:

A:Description: protein degradation

C:Superfamily: ubiquitin; ubiquitin homology

C:Keywords: nucleus; polyprotein; protein degradation

F:1-76/Product: ubiquitin #status predicted <UBI>

F:1-76/Domain: ubiquitin homology <UBH>

Query Match 35.5%; Score 38; DB 2; Length 76;  
Best Local Similarity 42.9%; Pred. No. 52;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHIDSOKKAI 21  
Db 10 GKTITLEVESSDTIDNVKAKI 30

## RESULT 7

S19799  
ubiquitin - potato

C:Species: Solanum tuberosum (potato)

C:Date: 30-Jun-1992 #sequence\_revision 03-Nov-2000 #text\_change 03-Nov-2000

C:Accession: S19799; S19801; S19800

R:Belknap, W.  
submitted to the EMBL Data Library, January 1992

A:Reference number: S19799

A:Accession: S19799  
A:Molecule type: mRNA

A:Residues: 1-56 <BEL1>  
A:Cross-References: EMBL:211671; NID:g21603; PIDN:CAA77737.1; PID:g21604  
A:Accession: S19801  
A:Molecule type: mRNA  
A:Residues: 1-48 <BEL2>  
A:Cross-References: EMBL:211672; NID:g21607; PIDN:CAA77738.1; PID:g21608  
A:Accession: S19800  
A:Molecule type: mRNA  
A:Residues: 37-77 <BEL3>  
A:Cross-References: EMBL:211673; NID:g21605; PIDN:CAA77739.1; PID:g21606  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: protein degradation  
F:1-76/Domain: ubiquitin homology <UB1>

Query Match 35.5%; Score 38; DB 2; Length 77;  
Best Local Similarity 42.9%; Pred. No. 53;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHDSOKKAI 21  
| : | | | | : | | : | |  
DB 10 GKTTLEVESDPTDVKAKI 30

RESULT 8  
S16263  
ubiquitin precursor - wheat (fragment)  
C:Species: Triticum aestivum (common wheat)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Jul-1999  
C:Accession: S16263  
R:Joshi, C.P.; Weng, J.; Nguyen, H.T.  
Plant Mol. Biol. 16, 907-908, 1991  
A>Title: Wheat ubiquitin gene exhibits a conserved protein coding region and a diverged  
A:Reference number: S16263; MUID:91316227  
A:Accession: S16263  
A:Molecule type: mRNA  
A:Residues: 1-77 <JOS>  
A:Cross-References: EMBL:X56601; NID:g21899; PIDN:CAA39938.1; PID:g21900  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: polypeptide  
F:1-76/Product: ubiquitin #status predicted <MAT>  
F:1-76/Domain: ubiquitin homology <OBH>

Query Match 35.5%; Score 38; DB 2; Length 77;  
Best Local Similarity 42.9%; Pred. No. 53;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHDSOKKAI 21  
| : | | | | : | | : | |  
DB 10 GKTTLEVESDPTDVKAKI 30

RESULT 9  
B96034  
degenerate transposase (orf1) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: B96034  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; Ee, R.; Leblanc, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: B96034  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <YUR>  
A:Cross-References: GB:AE007317; PIDN:AA00103.1; PID:g15458942; GSPDB:GN00174  
C:Genetics: IS1381-truncation  
A:Gene: IS1381-truncation

Query Match 35.5%; Score 38; DB 2; Length 79;

Best Local Similarity 50.0%; Pred. No. 54;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 6 VEPGSOHDSOKK 19  
| : | | | | : | | : | |  
DB 26 VEIPOSFDSKRR 39

RESULT 10  
F87520  
hypothetical protein CC2191 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: F87520  
R:Merz, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.;  
Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87520  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-80 <STO>  
A:Cross-References: GB:AE005673; NID:g13423692; PIDN:AAK24162.1; GSPDB:GN00148  
C:Genetics: CC2191  
A:Gene: CC2191

Query Match 35.5%; Score 38; DB 2; Length 80;  
Best Local Similarity 28.6%; Pred. No. 55;  
Matches 6; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHDSOKKAI 21  
| : | | | | : | | : | |  
DB 24 GDTFEGISPAHVEVERSYI 44

RESULT 11  
UOSY  
ubiquitin precursor - soybean (fragment)  
N:Alternate names: polyubiquitin  
C:Species: glycine max (soybean)  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Aug-1996  
C:Accession: S02222  
R:Portin, M.G.; Purohit, S.K.; Verma, D.P.S.  
Nucleic Acids Res. 16, 11377, 1988  
A>Title: The primary structure of soybean (glycine max) ubiquitin is identical to olt  
A:Reference number: S02222; MUID:89083578  
A:Accession: S02222  
A:Molecule type: mRNA  
A:Residues: 1-88 <FOR>  
A:Cross-References: EMBL:X13251  
C:Superfamily: ubiquitin; ubiquitin homology  
C:Keywords: polypeptide; protein degradation  
F:1-11/Product: ubiquitin (fragment) #status predicted <UB1>  
F:12-87/Product: ubiquitin #status predicted <UB2>  
F:12-87/Domain: ubiquitin homology <UBH>

Query Match 35.5%; Score 38; DB 1; Length 88;  
Best Local Similarity 42.9%; Pred. No. 61;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHDSOKKAI 21  
| : | | | | : | | : | |  
DB 21 GKTTLEVESDPTDVKAKI 41

RESULT 12  
S12161  
polyubiquitin - large-leaved lupine (fragment)  
C:Species: lupinus polyphyllus (large-leaved lupine)  
C>Date: 21-Nov-1993 #sequence\_revision 17-Jul-1998 #text\_change 23-Jul-1999  
C:Accession: S12161



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:23:28 : Search time 7.57377 Seconds  
(without alignments)  
107.359 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GEFQVEVPGSQHDSQKAI 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 12261

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	43	40.2	76	1	UBIQ_DICDI
2	40	37.4	76	1	P08618 dictyosteli
3	39	36.4	76	1	UBIQ_NEUCR
4	38	35.5	76	1	UBIQ_YEAST
5	38	35.5	76	1	UBIQ_COPCO
6	36	33.6	76	1	UBIQ_SOYBN
7	36	33.6	76	1	UBIQ_ACACA
8	36	33.6	77	1	UBIQ_PHYIN
9	34	31.8	53	1	ACP_SYNY3
10	34	31.8	74	1	YB80_ARCFU
11	34	31.8	76	1	CSRA_BACSU
12	34	31.8	76	1	UBIQ_ACECT
13	34	31.8	76	1	UBIQ_AGLNE
14	34	31.8	76	1	UBIQ_CAELT
15	34	31.8	76	1	UBIQ_CHLRE
16	34	31.8	76	1	UBIQ_TRYBB
17	34	31.8	98	1	UBIQ_TRYCR
18	34	31.8	99	1	R121_HALNI
19	33	30.8	45	1	VHSB_BP77
20	33	30.8	76	1	TYBA_ORPCA
21	33	30.8	76	1	UBIQ_EUPED
22	33	30.8	76	1	UBIQ_STRPU
23	33	30.8	81	1	ACP_GUTTH
24	32	29.9	43	1	TYB0_TORNA
25	32	29.9	75	1	CSRA_BACHD
26	32	29.9	76	1	UBIQ_DROME
27	32	29.9	76	1	UBIQ_HUMAN
28	32	29.9	78	1	UBIQ_LEIMA
29	32	29.9	97	1	ACP_HELP1
30	32	29.9	97	1	NB2M_BOVIN
31	32	29.9	97	1	PLAS_ORYZA
32	31.5	29.4	90	1	Y121_CHLPN
33	31	29.0	66	1	Z600_DROME
					COPP_HELP1
					Q9zm70 helicobacte

34	31	29.0	66	1	YSDA_ECOLI	P58096 escherichia
35	31	29.0	76	1	UBIQ_TETPY	P20685 tetrahymena
36	31	29.0	78	1	ACP_HELPY	P56464 helicobacte
37	31	29.0	80	1	YND2_MYCTU	Q50663 mycobacteri
38	31	29.0	89	1	SLTB_BP933	P09386 bacterioph
39	31	29.0	91	1	DEFW_MOUSE	P50715 mus musculu
40	31	29.0	91	1	DEFX_MOUSE	Q64263 mus musculu
41	31	29.0	91	1	DEFL_MOUSE	P17534 mus musculu
42	31	29.0	93	1	DEF1_MOUSE	P11477 mus musculu
43	31	29.0	93	1	FER2_EQUJE	P00236 equisetum t
44	31	29.0	94	1	Y761_SYNY3	P58236 synchocyst
45	31	29.0	98	1	FER1_ANASP	P06543 anabaena sp

## ALIGNMENTS

RESULT 1  
UBIQ\_DICDI STANDARD; PRT; 76 AA.  
AC P08618;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1991 (Rel. 20, Last annotation update)  
DE Ubiquitin.  
OS Dictyostelium discoideum ( slime mold ).  
OC Eukaryota; Eukaryota; Dictyostelidia; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88152253; PubMed=2831095;  
RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,  
Gerisch G.;  
RT "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-  
terminal tail and identification of the protein using an anti-peptide  
antibody.";  
RT FEBS Lett. 229:273-278(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89352609; PubMed=2548604;  
RA Ohmachi T., Giorda R., Shaw D.R., Ennis H.L.;  
RT "Molecular organization of developmentally regulated Dictyostelium  
discoideum ubiquitin cDNAs.";  
RN Biochemistry 28:5226-5231(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87257921; PubMed=3037345;  
RA Giorda R., Ennis H.L.;  
RT "Structure of two developmentally regulated Dictyostelium discoideum  
ubiquitin genes.";  
RT Mol. Cell. Biol. 7:2097-2103(1987).  
RN [4]  
RP SEQUENCE OF 13-76 FROM N.A.  
RC STRAIN=AX2;  
RA Westphal M., Mueller-Taubenberger A., Noegel A., Gerisch G.;  
RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
Dictyostelium discoideum.";  
RN FEBS Lett. 209:92-96(1986).  
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
CELLULAR PROTEINS. "THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
BIOGENESIS."  
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
CC -!- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
CURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).  
 -----  
 CC EMBL: X07210; CAA30183.1; ALT\_TERM.  
 DR EMBL: M19666; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19491; AAA33269.1; ALT\_TERM.  
 DR EMBL: M19492; AAA33270.1; ALT\_TERM.  
 DR EMBL: M23746; AAA33262.1; ALT\_TERM.  
 DR EMBL: M23749; AAA33263.1; ALT\_TERM.  
 DR EMBL: M23750; AAA33264.1; ALT\_TERM.  
 DR EMBL: M23751; AAA33265.1; ALT\_TERM.  
 DR EMBL: M23752; AAA33266.1; ALT\_TERM.  
 DR EMBL: M23753; AAA33267.1; ALT\_TERM.  
 DR EMBL: M23754; AAA33268.1; ALT\_TERM.  
 DR EMBL: X04702; CAA28408.1; ALT\_TERM.  
 DR PIR: S00357; UDOR.  
 DR PIR: E34080; UDOR.  
 DR PIR: A27806; A27806.  
 DR PIR: B27806; B27806.  
 DR PIR: A34080; A34080.  
 DR PIR: B34080; B34080.  
 DR PIR: C34080; C34080.  
 DR PIR: D34080; D34080.  
 DR HSSP: P02248; IUBI.  
 DR DictyDb: DD05001; ubqA.  
 DR InterPro: IPR000628; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin.  
 DR PRINTS: PR00348; Ubiquitin.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; Ubiquitin\_1; 1.  
 DR PROSITE: PS50053; Ubiquitin\_2; 1.  
 KW Nuclear protein; Polypeptide.  
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76 76 MULTIBIQUITIN ADDUCTS.  
 FT VARIANT 11 11 CONJUGATION TO ACCEPTOR PROTEINS.  
 FT SEQUENCE 76 AA; 8538 MW; 6427383968EAB84 CRC64;  
 SO Query Match 40.2%; Score 43; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 3.2;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
 Oy 1 GETFOVEYPSQHIDSOKKAI 21  
 Db 10 GKTITLEVSGSDNIENYKAKI 30  
 RESULT 2  
 UBIO\_NEUCR STANDARD; PRT; 76 AA.  
 AC P13117;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ubiquitin.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ST. LAWRENCE 74 / SL 74 / ORS 6A;  
 RX MEDLINE=8936647; PubMed=2549509;  
 RA Tacciolli G.E., Grotebold E., Aisemberg G.O., Judewicz N.D.;  
 RT "Ubiquitin expression in Neurospora crassa: cloning and sequencing of  
 a polyubiquitin gene.";  
 RL Nucleic Acids Res. 17:6153-6165(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ST. LAWRENCE 74 / SL 74 / ORS 6A;  
 RX MEDLINE=9133720; PubMed=1650731;  
 RA Tacciolli G.E., Grotebold E., Aisemberg G.O., Judewicz D.N.;  
 RT "The cDNA sequence and expression of an ubiquitin-tail gene fusion in  
 Neurospora crassa.";

RL Gene 102:133-137(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A;  
 RX MEDLINE=94374698; PubMed=8086539;  
 RA Taraneh K.A., Annula K.R., Free S.J.;  
 RT "The isolation and characterization of a Neurospora crassa gene  
 (ubi::crp-6) encoding a ubiquitin-40S ribosomal protein fusion  
 protein.";  
 RL Gene 147:137-140(1994).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 PRECURSOR WITH 4 EXACT HEAD TO TAIL REPEATS. THERE IS A FINAL  
 AMINO-ACID (GIN) AFTER THE LAST REPEAT. SOME UBIQUITIN GENES  
 CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
 BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -  
 THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
 USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
 MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
 ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.lsb-sib.ch/announce/>  
 or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC EMBL: X13140; CAA31530.1; ALT\_TERM.  
 DR EMBL: U01220; AAA56880.1; ALT\_TERM.  
 DR EMBL: U01221; AAA03351.1; ALT\_TERM.  
 DR PIR: S05323; UQNC.  
 DR HSSP: P02248; IUBI.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin.  
 DR PRINTS: PR00348; Ubiquitin.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; Ubiquitin\_1; 1.  
 DR PROSITE: PS50053; Ubiquitin\_2; 1.  
 KW Nuclear protein; Polypeptide.  
 FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN  
 FT BINDING 76 76 MULTIBIQUITIN ADDUCTS.  
 FT SEQUENCE 76 AA; 8598 MW; 994480FE7D38403E CRC64;  
 SO Query Match 37.4%; Score 40; DB 1; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 9.7;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 Oy 1 GETFOVEYPSQHIDSOKKAI 21  
 Db 10 GKTITLEVSGSDTIIDVKKKI 30  
 RESULT 3  
 UBIO\_YEAST STANDARD; PRT; 76 AA.  
 AC P04838;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Ubiquitin.  
 GN (UBI1 OR RPL40A OR YIL148W) AND (UBI2 OR RPL40B OR YKR094C) AND  
 GN (UBI3 OR RPS31 OR YLR167W OR L9470.14) AND (UBI4 OR SCD2 OR YIL039C).  
 OS Saccharomyces cerevisiae (Baker's yeast),  
 OS Candida albicans (Yeast), and  
 OS Cryptococcus neoformans (Filobasidiella neoformans).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932, 5476, 5207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.cerevisiae;

RX MEDLINE=85061630; PubMed=6095120;  
 RA Oezkaynak E., Finley D., Varshavsky A.;  
 RT "The yeast ubiquitin gene: head-to-tail repeats encoding a  
 RT polyubiquitin precursor protein.";  
 RL Nature 312:663-666(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (UBI1; UBI2; UBI3 AND UBI4).  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=87275838; PubMed=3038523;  
 RA Oezkaynak E., Finley D., Solomon M.J., Varshavsky A.;  
 RT "The yeast ubiquitin genes: a family of natural gene fusions.";  
 RL EMO J. 6:1429-1439(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A. (UBI2).  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=94205265; PubMed=8154186;  
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
 RT cerevisiae chromosome XI contains the UBI2 and MPL genes and three  
 RT new open reading frames.";  
 RL Yeast 9:1349-1354(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A. (UBI1).  
 RC SPECIES=S.cerevisiae; STRAIN=S288C / AB972;  
 RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,  
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (UBI3).  
 RC SPECIES=S.cerevisiae; STRAIN=S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favell A., Fallon L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba D., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston J., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Tatch A., Trevasakis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A. (UBI4).  
 RC SPECIES=C.albicans;  
 RA Bailey D.A., Gow N.A.R., Brown A.J.P.;  
 RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A. (UBI1).  
 RC SPECIES=C.albicans; STRAIN=ATCC 26555;  
 RA Martinez J.P.;  
 RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C.neoformans; STRAIN=ATCC 6352;  
 RX MEDLINE=96369720; PubMed=7642124;  
 RA Spltzer E.D., Spltzer S.G.;  
 RT "Structure of the ubiquitin-encoding genes of Cryptococcus  
 RT neoformans.";  
 RL Gene 161:113-117(1995).  
 RN [9]  
 RP MUTAGENESIS OF LYS-29; LYS-48 AND GLY-76.  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=95340540; PubMed=7615550;  
 RA Johnson E.S., Ma P.C.M., Ota M., Varshavsky A.;  
 RT "A proteolytic pathway that recognizes ubiquitin as a degradation  
 RT signal.";  
 RL J Biol. Chem. 270:17442-17456(1995).  
 RN [10]  
 RP MUTAGENESIS OF LYSINE RESIDUES.  
 RC SPECIES=S.cerevisiae;  
 RX MEDLINE=95166210; PubMed=7862120;  
 FT

RA Spence J., Sadis S., Haas A.L., Finley D.;  
 RT "A ubiquitin mutant with specific defects in DNA repair and  
 RT multibubination.";  
 RL Mol. Cell. Biol. 15:1265-1273(1995).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS. THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS AND DNA REPAIR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBI1 AND UBI2 ARE SYNTHESIZED AS A POLYPEPTIDE WITH  
 CC ONE COPY OF UBIQUITIN FUSED TO RIBOSOMAL PROTEIN 537. UBI3 IS A  
 CC POLYPEPTIDE WITH ONE COPY OF UBIQUITIN FUSED TO ANOTHER RIBOSOMAL  
 CC PROTEIN.  
 CC -1- MISCELLANEOUS: IN S.CEREVISIAE UBI4 IS A POLYPEPTIDE CONTAINING 5  
 CC EXACT HEAD TO TAIL REPEATS OF UBIQUITIN, THERE IS A FINAL  
 CC AMINO-ACID (ASN) AFTER THE LAST REPEAT.  
 CC -1- MISCELLANEOUS: IN C.ALBICANS UBI4 IS A POLYPEPTIDE CONTAINING 3  
 CC EXACT HEAD TO TAIL REPEATS OF UBIQUITIN, THERE IS A FINAL  
 CC AMINO-ACID (PHE) AFTER THE LAST REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 DR EMBL: X01473; CAA25706.1; ALT\_TERM.  
 DR EMBL: X01474; CAA25706.1; ALT\_SEQ.  
 DR EMBL: X05728; CAA29195.1; ALT\_TERM.  
 DR EMBL: X05729; CAA29196.1; ALT\_TERM.  
 DR EMBL: X05730; CAA29197.1; ALT\_TERM.  
 DR EMBL: X05731; CAA29198.1; ALT\_TERM.  
 DR EMBL: X73541; CAA51949.1; ALT\_TERM.  
 DR EMBL: X28319; CAA82173.1; ALT\_TERM.  
 DR EMBL: U17246; AAB67466.1; ALT\_TERM.  
 DR EMBL: U54197; CAA80901.1; ALT\_TERM.  
 DR EMBL: Z38059; CAA86130.1; ALT\_TERM.  
 DR EMBL: U32627; AAB48686.1; ALT\_TERM.  
 DR EMBL: U16992; AAB2979.1; ALT\_TERM.  
 DR PIR: A22696; UOBY.  
 DR PIR: C29456; UOBYR7.  
 DR PIR: A29456; A29456.  
 DR PIR: B28456; B28456.  
 DR PIR: D29456; D29456.  
 DR HSP: P02248; UBI1.  
 DR SWISS-2DPAGE; P04838; YEAST.  
 DR SGD: S0001410; RPL40A.  
 DR SGD: S0001802; RPL40B.  
 DR SGD: S0004157; RPS31.  
 DR SGD: S0003962; UBI4.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00289; UBIQUITIN\_1; 1.  
 DR PROSITE: PS0053; UBIQUITIN\_2; 1.  
 KW Nuclear protein; Polyprotein; Multigene family; DNA repair.  
 FT SITE 29 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 48 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 63 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT SITE 76 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT BINDING 76 76 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 29 29 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 48 48 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT MUTAGEN 63 63 INVOLVED IN UB-UB LINKAGES OF MULTI-  
 FT FUNCTION.





RT tobacco.";  
 RL [7] J. Biol. Chem. 265:12486-12493(1990).  
 RP SEQUENCE FROM N.A. (UBQ13).  
 RC SPECIES-A.thaliana;  
 RX MEDLINE-93177218; PubMed-8382544;  
 RA Sun C.W., Callis J.;  
 RT "Recent stable insertion of mitochondrial DNA into an Arabidopsis  
 RL plant cell 5:97-107(1993).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-H.vulgaris; STRAIN-CV. BOMI;  
 RX MEDLINE-86274685; PubMed-2426105;  
 RA Gausing K., Barkardottir R.;  
 RT "Structure and expression of ubiquitin genes in higher plants.";  
 RL Eur. J. Biochem. 158:57-62(1986).  
 RN [9]  
 RP SEQUENCE FROM N.A. (MUB1 AND MUB2).  
 RC SPECIES-H.vulgaris; STRAIN-CV. BOMI;  
 RX MEDLINE-91078635; PubMed-1701748;  
 RA Gausing K., Jensen C.B.;  
 RT "Two ubiquitin-long-tail fusion genes arranged as closely spaced  
 RL direct repeats in barley.";  
 RN Gene 94:165-171(1990).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-T.aestivum; STRAIN-CV. MUSTANG; TISSUE-Leaf;  
 RX MEDLINE-91316227; PubMed-1650258;  
 RA Joshi C.P., Weng J., Nguyen H.T.;  
 RT "Wheat ubiquitin gene exhibits a conserved protein coding region and  
 RL a diverged 3' non-coding region.";  
 RN Plant Mol. Biol. 16:907-908(1991).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-H.annuus; STRAIN-CV. HA401B / CARGILL;  
 RX MEDLINE-89183614; PubMed-2538802;  
 RA Binet M.N., Steilmetz A., Tessier L.H.;  
 RT "The primary structure of sunflower (Helianthus annuus) ubiquitin.";  
 RL Nucleic Acids Res. 17:2119-2119(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Maize;  
 RA Christensen A.H., Quail P.H.;  
 RT "Sequence analysis and transcriptional regulation by heat shock of  
 RL polyubiquitin transcripts from maize.";  
 RN Plant Mol. Biol. 12:619-632(1989).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Maize;  
 RX MEDLINE-92084112; PubMed-1660830;  
 RA Chen K., Rubenstein I.;  
 RT "Characterization of the structure and transcription of an ubiquitin  
 RL fusion gene from maize.";  
 RN Gene 107:205-212(1991).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-L.polyphyllus;  
 RX MEDLINE-91057134; PubMed-2173830;  
 RA Perrey R., Warskulat U., Wink M.;  
 RT "Molecular cloning of a cDNA for the ubiquitin gene of Lupinus  
 RL polyphyllus.";  
 RN Nucleic Acids Res. 18:6428-6428(1990).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-L.albus;  
 RX MEDLINE-94156199; PubMed-8112604;  
 RA Jacinto A., Neves A.M., Vassilevskaia T.D., Ricardo C.P.,  
 RA Rodrigues-Pousada C.;  
 RT "Cloning and characterization of two ubiquitin::79-amino-acid  
 RL extension protein-encoding fusion genes from Lupinus albus.";  
 RN Gene 139:201-205(1994).  
 OX [16]

RP SEQUENCE FROM N.A.  
 RC SPECIES-L.usitatissimum;  
 RX MEDLINE-91216447; PubMed-1850710;  
 RA Agarwal M.L., Cullis C.A.;  
 RT "The ubiquitin-encoding multigene family of flax, Linum  
 RL usitatissimum.";  
 RN Gene 99:69-75(1991).  
 RN [17]  
 RP SEQUENCE FROM N.A. (UBI3).  
 RC SPECIES-L.esculentum; TISSUE-Root;  
 RX MEDLINE-92032784; PubMed-1657246;  
 RA Hoffman N.E., Ko K., Milkowski D., Pichersky E.;  
 RT "Isolation and characterization of tomato cDNA and genomic clones  
 RL encoding the ubiquitin gene ubi3.";  
 RN Plant Mol. Biol. 17:1189-1201(1991).  
 RN [18]  
 RP SEQUENCE FROM N.A. (UBI3).  
 RC SPECIES-S.tuberosum; STRAIN-CV. LEWNI RUSSET; TISSUE-Tuber;  
 RX MEDLINE-93004476; PubMed-1327270;  
 RA Garbarno J.E., Rockhold D.R., Belknap W.R.;  
 RT "Expression of stress-responsive ubiquitin genes in potato tubers.";  
 RL Plant Mol. Biol. 20:235-244(1992).  
 RN [19]  
 RP SEQUENCE FROM N.A. (UBI3).  
 RC SPECIES-S.tuberosum; STRAIN-CV. LEWNI RUSSET; TISSUE-Tuber;  
 RX MEDLINE-94154225; PubMed-8111011;  
 RA Garbarno J.E., Belknap W.R.;  
 RT "Isolation of a ubiquitin-ribosomal protein gene (ubi3) from potato  
 RL and expression of its promoter in transgenic plants.";  
 RN Plant Mol. Biol. 24:119-127(1994).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-N.sylvestris; TISSUE-Leaf;  
 RX MEDLINE-9309242; PubMed-1281439;  
 RA Genschik P., Parmentier Y., Durr A., Marbach J., Criqui M.C.,  
 RA Jamet E., Fleck J.;  
 RT "Ubiquitin genes are differentially regulated in protoplast-derived  
 RL cultures of Nicotiana sylvestris and in response to various  
 RT stresses.";  
 RL Plant Mol. Biol. 20:897-910(1992).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-A.officinalis; STRAIN-CV. LIMBRAS 10; TISSUE-Spear;  
 RA Davies K.M., King G.A.;  
 RT "Isolation and characterization of Asparagus officinalis L. cDNA  
 RL clones encoding two forms of ubiquitin mRNA.";  
 RN New Zealand J. Crop Hort. Sci. 21:153-159(1993).  
 RN [22]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-O.sativa;  
 RX MEDLINE-93271454; PubMed-8388748;  
 RA Nishi P., Hashimoto H., Kidou S.I., Uchimaya H., Kato A.;  
 RA "Query Match 35.5%; Score 38; DB 1; Length 76;  
 Best local Similarity 42.9%; Pred. No. 20;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 GETFOVEVPGSHIDSOKKAI 21  
 Db 10 GKTTLEVESSDTIDWYKAKI 30  
 ID UBIO\_ACACA STANDARD; PRT; 76 AA.  
 AC P49634;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE ubiquitin.  
 OS Acanthamoeba castellanii (Amoeba).  
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 OX NCBI\_TaxID=5755;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94250685; PubMed=8193157;
RA Am R.S., Henney H.R.;
RT "An Acanthamoeba ubiquitin-fusion protein; cDNA and deduced protein
sequence."
RL Biochem. Biophys. Acta 1218:109-111(1994).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED WITH RIBOSOMAL CEP2 AS
CC ITS C-TERMINAL EXTENSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75628; CA53293.1; ALT_TERM.
DR HSSP: P02248; IUBI.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR SMART: SM00213; UBO; 1.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS50053; UBIQUITIN_2; 1.
KW Nuclear protein; Polypeptide.
FT SITE 48
FT NEEDED FOR BRANCHED-CHAIN
FT MULTIBIQUITIN ADDUCTS.
FT BINDING 76 76
FT SEQUENCE 76 AA; 8596 MW; D84480E07D1E9B52 CRC64;
SQ
Query Match 33.68; Score 36; DB 1; Length 76;
Best Local Similarity 38.18; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 10 KTTTLEVESDTEYNKQKI 30

```

```

CC AMINO-ACID (ASN) AFTER THE LAST REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55717; CA53250.1; ALT_TERM.
DR PIR: S17740; UOJN1.
DR HSSP: P02248; IUBI.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR PRINTS: PR00348; UBIQUITIN.
DR SMART: SM00213; UBO; 1.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS50053; UBIQUITIN_2; 1.
KW Nuclear protein; Polypeptide.
FT SITE 48
FT NEEDED FOR BRANCHED-CHAIN
FT MULTIBIQUITIN ADDUCTS.
FT BINDING 76 76
FT SEQUENCE 76 AA; 8580 MW; 98F084FE7F3A456F CRC64;
SQ
Query Match 33.68; Score 36; DB 1; Length 76;
Best Local Similarity 38.18; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
OY 1 GETFOVEVPGSOHIDSOKKAI 21
Db 10 KTTTLEVESDTEYNKQKI 30

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90905; BAA17312.1; -  
 DR PIR: S13820; S13820.  
 DR HSSP: P02901; IACP.  
 DR InterPro: IPR003331; Acyl\_carrier.  
 DR InterPro: IPR003880; Phosphopant\_attach.  
 DR Pfam: PF000550; pp-binding; 1.  
 DR ProDom: PD000887; Acyl\_carrier; 1.  
 DR ProSITE: PS00012; PHOSPHOPANTHETHEINE; 1.  
 DR ProSITE: PS50075; ACP\_DOMAIN; 1.  
 DR Fatty acid biosynthesis; Phosphopantetheine; Complete proteome.  
 KW BINDING 38 38 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT CONFLICT 2 2 N -> D (IN REF. 2).  
 FT CONFLICT 23 23 K -> G (IN REF. 2).  
 SQ SEQUENCE 77 AA; 8590 MW; FDDDED4BDFDEBDB53 CRC64;  
 Query Match 33.6%; Score 36; DB 1; Length 77;  
 Best Local Similarity 36.4%; Pred. No. 43;  
 Matches 8; Conservative 6; Mismatches 6; Indels 2; Gaps 1;  
 QY 2 ETPFQVEVPG--SQHDSQKAI 21  
 DB 50 EEFDEIPDEVAETIDTVGKAV 71  
 ID YE80\_ARCFU STANDARD; PRT; 53 AA.  
 AC 028792;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AF1480.  
 GN AF1480.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Ritschman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kikness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uutterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE001000; AAB89767.1; -  
 DR TIGR: AF1480; -

KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 53 AA; 5955 MW; B485651A92197239 CRC64;  
 Query Match 31.8%; Score 34; DB 1; Length 53;  
 Best Local Similarity 35.6%; Pred. No. 60;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GETFQVEVP 9  
 DB 42 GDTFQLOAP 50  
 ID CSRA\_BACSU STANDARD; PRT; 74 AA.  
 AC P33911;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Carbon storage regulator homolog.  
 GN CSRA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89255067; PubMed=2498284;  
 RA Mirel D.B., Chamberlin M.J.;  
 RT "The Bacillus subtilis flagellin gene (hag) is transcribed by the  
 RT sigma 28 form of RNA polymerase."  
 RL J. Bacteriol. 171:3095-3101(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Soldo B., Lazarevic V., Mauei C., Karamata D.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP IDENTIFICATION.  
 RA Robinson K.;  
 RL Unpublished observations (SEP-1993).  
 CC -I- FUNCTION: COULD ACCELERATE THE DEGRADATION OF SOME GENES  
 CC TRANSCRIPTS POTENTIALLY THROUGH SELECTIVE RNA BINDING  
 CC (BY SIMILARITY)  
 CC -I- SIMILARITY: BELONGS TO THE CSRA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M26948; -; NOT ANNOTATED\_CDS.  
 DR EMBL: U56901; AAC44950.1; -  
 DR EMBL: Z99122; CAB15554.1; -  
 DR Subtilisin; Bg10654; csra.  
 DR InterPro: IPR003751; CSRA.  
 DR Pfam: PF02599; CSRA; 1.  
 KW RNA-binding; Complete proteome.  
 FT DOMAIN 11 45 KH.  
 SQ SEQUENCE 74 AA; 8136 MW; 10E82DA62D3174B5 CRC64;  
 Query Match 31.8%; Score 34; DB 1; Length 74;  
 Best Local Similarity 28.6%; Pred. No. 86;  
 Matches 6; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 GETFQVEVPGSQHDSQKAI 21  
 DB 27 GQVVKIGIDAPKRIHRIKREI 47



```

RP SEQUENCE OF 1-37 FROM N.A.
RC SPECIES=C.elegans;
RX MEDLINE=86273146; Pubmed-2839490;
RA Graham R.W., van Doren K., Bektess S., Candido E.P.M.;
RT "Maturation of the major ubiquitin gene transcript in Ctenophoraditits
RT elegans involves the acquisition of a trans-spliced leader.";
RL J. Biol. Chem. 263:10415-10419(1988).
RN [4]
RP SEQUENCE FROM N.A. (UBQ-2).
RC SPECIES=C.elegans; STRAIN=BRISTOL N2;
RX MEDLINE=96018742; Pubmed-7556908;
RA Jones D., Candido P.M.;
RT "A portable regulatory element directs specific expression of the
RT Ctenophoraditits elegans ubiquitin gene ubq-2 in the somatic gonad.";
RL Dev. Biol. 171:60-72(1995).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=G.cyclopium;
RX MEDLINE=94110294; Pubmed-8282761;
RA Pfeiffer K., Frank K., Schroeder H.C., Gamulin V., Rinkevich B.,
RA Batel R., Mueller I.M., Mueller E.G.;
RT "Cloning of the polyubiquitin cDNA from the marine sponge Geodia
RT cyclops and its preferential expression during reaggregation of
RT cells.";
RL J. Cell. Sci. 106:545-554(1993).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN
CC PRECURSOR WITH 11 EXACT HEAD TO TAIL REPEATS. THERE ARE TWO
CC ADDITIONAL AMINO-ACIDS (ASP-TLE) AFTER THE LAST REPEAT.
CC -----
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M23433; AAA28154.1; ALT_TERM.
DR EMBL, U23172; AAC46525.1; ALT_TERM.
DR EMBL, M21321; AAA28153.1; -.
DR EMBL, L31492; AAC37252.1; ALT_TERM.
DR EMBL, X70917; CAA50268.1; ALT_TERM.
DR PIR, A30126; A30126.
DR PIR, A28804; A28804.
DR HSSP, P02248; IOBI.
DR Wormpep; F25H5.4; CE01921.
DR InterPro; IPRO00626; Ubiquitin.
DR Pfam; PF00240; ubiquitin.1.
DR PRINTS; PR00348; UBIQUITIN.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
KW Nuclear protein; Polyprotein.
FT SITE 48
FT 48
FT NECESSARY FOR BRANCHED-CHAIN
FT UBIQUITIN ADDUCTS.
FT CONUNGATION TO ACCEPTOR PROTEINS.
FT BINDING 76
FT 76
SQ SEQUENCE 76 AA; 8539 MW; C43B24396ECC9B52 CRC64;
Query Match 31.88; Score 34; DB 1; Length 76;
Best Local Similarity 38.18; Pred. No. 89;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
QY 1 GETFOVEPGSOHIDSOKKAI 21
|:|:|:|:|:|:|:|:|:|
Db 10 GKTITLEVADSLTEINVKAKI 30
RESULT 14

```

```

UBIQ_CHLRE          STANDARD:      PRT:       76 AA.
ID   UBIQ_CHLRE     PI4624;
AC   UBIQ_CHLRE     01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Ubiquitin.
GN   UBI1 AND UBI3.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A. (UBI1).
RC   STRAIN=2137;
RA   MEDLINE=90045969; PubMed=2554258;
RX   Callis J., Pollmann L., Shanklin J., Wettern M.; Vlerstra R.;
RT   "Sequence of a cDNA from Chlamydomonas reinhardtii encoding a
RL   ubiquitin 52 amino acid extension protein."
RN   [2]
RP   SEQUENCE FROM N.A. (UBI3).
RC   MEDLINE=92037644; PubMed=1657605;
RX   Pollmann L., von Kampen J., Wettern M.;
RT   "Ubiquitin in a lower plant. Characterization of ubiquitin-encoding
RN   DNA and RNA from Chlamydomonas reinhardtii."
RL   Eur. J. Biochem. 202:197-204(1991).
CC   -1 FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF
CC   CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE
CC   REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME
CC   BIOGENESIS.
CC   -1 SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC   -1 MISCELLANEOUS: UBI1 IS SYNTHESIZED AS A POLYPEPTIDE WITH ONE COPY
CC   OF UBQUITIN FUSED TO A RIBOSOMAL PROTEIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR   EMBL; X15427; CAA33466.1; ALT_TERM.
DR   EMBL; X60826; CAA43216.1; ALT_TERM.
DR   PIR; S06598; UOKM.
DR   HSSP; P02248; IUBI.
DR   InterPro; IPR000626; Ubiquitin.
DR   Pfam; PF00240; ubiquitin_1.
DR   PRINTS; PR00348; UBIQUITIN.
DR   SMART; SM00213; UBO; 1.
DR   PROSITE; PS00293; UBIQUITIN_1; 1.
DR   PROSITE; PS50053; UBIQUITIN_2; 1.
DR   TM   Nuclear protein; Polyprotein.
FT   SITE           48                               48
FT   BINDING        76 AA; 8539 MW; C42936277D1E9B52 CRC64;
FT   FT            76                               76
SO   SEQUENCE

Query Match      31.8%; Score 34; DB 1; Length 76;
Best Local Similarity 38.1%; Pred. No. 89;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY   1 GETFOVEPPGSHIDSOKKAI 21
    | : | | | : | | |
DB   10 KXTITLEVESDTEIENVRAKI 30

RESULT 15
UBIO_TRYBB
ID   UBIO_TRYBB     STANDARD:      PRT:       76 AA.
AC   P15174;
DT   01-APR-1990 (Rel. 14, Created)

```

Search completed: October 24, 2002, 15:35:32  
 Job time : 7.57377 secs

```

DT 01-APR-1990 (Rel. 14, last sequence update)
DT 01-FEB-1995 (Rel. 31, last annotation update)
DE Ubiquitin.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxId=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90136711; PubMed=2559328;
RA Wong S., Campbell D.A.;
RT "A polyubiquitin gene from Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 37:147-150(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=427;
RC MEDLINE=91088342; PubMed=2175891;
RA Wong S., Morales T.H., Campbell D.A.;
RT "Ubiquitin-EP52 fusion protein homologs from Trypanosoma brucei.";
RL Nucleic Acids Res. 18:7181-7181(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=427;
RC MEDLINE=93109304; PubMed=8380221;
RA Wong S., Morales T.H., Neigel J.E., Campbell D.A.;
RT "Genomic and transcriptional linkage of the genes for calmodulin, EF-hand 5 protein, and ubiquitin extension protein 52 in Trypanosoma brucei.";
RL Mol. Cell. Biol. 13:207-216(1993).
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN PRECURSOR WITH 5 OR 6 EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X14554; CAA32691.1; ALT_TERM.
DR EMBL: X54641; CAA38453.1; -.
DR EMBL: X54642; CAA38454.1; -.
DR EMBL: X56511; CAA39863.1; -.
DR EMBL: X56511; CAA39864.1; -.
DR PIR: S07998; UOQT.
DR PIR: S14197; S14197.
DR HSSP: P02248; IUBI.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; Ubiquitin.
DR PRINTS: PR00348; UBIQUITIN.
DR SMART: SM00213; UBO; 1.
DR PROSITE: PS00299; UBIQUITIN_1; 1.
DR PROSITE: PS50053; UBIQUITIN_2; 1.
KW Nuclear protein; Polypeptide.
FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN MULTIBIQUITIN ADDUCTS.
FT BINDING 76 76 CONJUGATION TO ACCEPTOR PROTEINS.
SQ SEQUENCE 76 AA; 8507 MW; C53759576F0C8E47 CRC64;
Query Match 31.88; Score 34; DB 1; Length 76;
Best Local Similarity 38.18; Pred. No. 89;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Caps 0;
OY 1 GETFOVEVPGSQHIDSOKKAI 21
DB 10 GKTALAEVASTITENYKAKI 30

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:24:04 ; Search time 24.4426 Seconds  
(without alignments)  
148.629 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEVPGSOHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 107123

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_oranelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_ricket:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	43.0	91	10	Q9SYF4
2	40	37.4	79	16	Q97QBI
3	40	37.4	79	16	Q97P24
4	40	37.4	91	10	Q9M4A4
5	39	36.4	45	5	Q36044
6	39	36.4	47	3	Q07188
7	39	36.4	62	5	Q36058
8	39	36.4	62	5	Q36043
9	39	36.4	77	11	Q9QYJ8
10	39	36.4	77	16	Q9AP16
11	38	35.5	38	10	Q9SAT2
12	38	35.5	52	10	Q42009
13	38	35.5	54	10	Q42001
14	38	35.5	72	10	Q9M646
15	38	35.5	76	10	Q94IM2
16	38	35.5	76	10	Q93XW4

17	38	35.5	77	10	Q41570
18	38	35.5	77	10	P93379
19	38	35.5	78	10	Q24122
20	38	35.5	80	16	Q9A6A3
21	38	35.5	92	10	Q42202
22	38	35.5	92	10	Q941N9
23	38	35.5	99	10	Q41431
24	38	35.5	100	10	Q9AY08
25	37	34.6	73	6	Q9GMA0
26	37	34.6	76	3	Q9P323
27	37	34.6	79	4	Q96HT2
28	37	34.6	83	7	Q19274
29	37	34.6	94	12	Q91F01
30	36	33.6	44	10	Q42153
31	36	33.6	77	2	Q9R340
32	36	33.6	83	12	Q99H19
33	36	33.6	91	16	Q9JRI4
34	35	32.7	41	2	Q9X920
35	35	32.7	65	10	Q9M4H3
36	35	32.7	68	2	Q9F2S1
37	35	32.7	73	5	Q25833
38	35	32.7	78	16	Q9A7P3
39	35	32.7	89	2	Q9RHN9
40	35	32.7	97	2	Q47597
41	35	32.7	99	8	Q37764
42	34.5	32.2	89	2	Q9AKD3
43	34	31.8	67	13	Q91B57
44	34	31.8	79	4	Q9NYS3
45	34	31.8	80	5	Q91W97

## ALIGNMENTS

RESULT 1					
Q9SYF4	PRELIMINARY;	PRT;	91 AA.		
AC	Q9SYF4:				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	F1511.6 PROTEIN.				
GN	F1511.6.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RA	Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,				
RA	Li J., Kremenetskaia I., Lueros J., Ngan I., Gonzalez A., Altati H.,				
RA	Aravajo R., Chao O., Conn L., Conway A.B., Dunn P., Hansen N.,				
RA	Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,				
RA	Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,				
RT	"Arabidopsis thaliana chromosome 1 BAC F1511 sequence."				
RT	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AC006577; AAD25769.1; -.				
DR	HSSP; P02248; IUBI.				
DR	InterPro; IPR000626; Ubiquitin.				
DR	Pfam; PF00240; ubiquitin; 2.				
SO	SEQUENCE 91 AA; 10142 MW; E8766823D6450267 CRC64;				
QY	1 GETFOVEVPGSOHIDSOKKAI 21				
DB	58 GKTFNLEVKSETIQVKNNI 78				
	Query Match	43.0%;	Score 46;	DB 10;	Length 91;
	Best Local Similarity	47.6%;	Pred. No. 3.8;		
	Matches 10;	Conservative 3;	Mismatches 8;	Indels 0;	Gaps 0;

<b>RESULT 2</b>					
ID	097OB1	PRELIMINARY;	PRT;	79 AA.	
AC	097OB1;				
DT	01-OCT-2001 (TREMBLrel. 18, Created)				
DR	01-OCT-2001 (TREMBLrel. 18, Last sequence update)				
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
GN	IS66 FAMILY ELEMENT, ORF1.				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
CC	Streptococcus.				
OX	NCBI_TaxID=1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=TIGR4;				
RX	MEDLINE=21357209; PubMed=11463916;				
RA	Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,				
RA	Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,				
RA	Durkin A.S., Gwinn M., Kolony J.F., Nelson W.C., Peterson J.D.,				
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,				
RA	Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,				
RA	McDonald L.A., Feldlyum T.V., Angiolini S., Dickinson T., Hickey E.K.,				
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., C.M.,				
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;				
RT	"Complete genome sequence of a virulent isolate of Streptococcus				
RT	pneumoniae.";				
RL	Science 293:498-506(2001).				
DR	EMBL: AE0074429; AAK75412.1; -.				
KW	TIGR: SP1314; -.				
KM	Complete proteome.				
SQ	SEQUENCE 79 AA; 9260 MW; B5319A03D3235FE2 CRC64;				
Query Match		37.4%;	Score 40;	DB 16;	Length 79;
Best Local Similarity		50.0%;	Pred. No. 33;		
Matches	7;	Conservative	4;	Mismatches	3;
				Indels	0;
				Gaps	0;
OY	6 VEVPQSQHDSQRK 19				
Db	26 VEIPQSRREDSSKR 39				
<b>RESULT 3</b>					
ID	097PZ4	PRELIMINARY;	PRT;	79 AA.	
AC	097PZ4;				
DT	01-OCT-2001 (TREMBLrel. 18, Created)				
DR	01-OCT-2001 (TREMBLrel. 18, Last sequence update)				
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
GN	IS66 FAMILY ELEMENT, ORF1.				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
CC	Streptococcus.				
OX	NCBI_TaxID=1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=TIGR4;				
RX	MEDLINE=21357209; PubMed=11463916;				
RA	Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,				
RA	Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,				
RA	Durkin A.S., Gwinn M., Kolony J.F., Nelson W.C., Peterson J.D.,				
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,				
RA	Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,				
RA	McDonald L.A., Feldlyum T.V., Angiolini S., Dickinson T., Hickey E.K.,				
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,				
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;				
RT	"Complete genome sequence of a virulent isolate of Streptococcus				
RT	pneumoniae.";				
RL	Science 293:498-506(2001).				
DR	EMBL: AE007441; AAK7537.1; -.				
KW	TIGR: SP1443; -.				
KM	Complete proteome.				

50	SEQUENCE	79 AA;	9294 MW;	553CFA03DB89FDD6A CRC64;
	Query Match	37.4%;	Score 40;	DB 16; Length 79;
	Best Local Similarity	50.0%;	Pred. No. 33;	
	Matches	7; Conservative	4; Mismatches	3; Indels 0; Gaps 0;
QY	6	VEVPGSOHIDSOKR 19		
		: : : : : : :		
Db	26	VEIPQSRFRDSSKR 39		
		: : : : : : :		
	RESULT 4			
	Q9M4A4			
ID	Q9M4A4	PRELIMINARY;	PRT;	91 AA.
AC	Q9M4A4;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PURATIVE RESISTANCE GENE HOMOLOGUE (FRAGMENT).			
GN	MRCR-4.			
OS	Cucumis melo (Muskmelon).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euphorbia; Cucurbitales; Cucurbitaceae; Cucumis.			
OX	NCBI_TaxID=3656;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LEAF;			
RA	Garcia-Mas J., van Leeuwen H., Monfort A., de Vicente M.C.,			
RA	Puligdomenech P., Arus P.,			
RT	"Cloning and mapping of resistance gene homologues in melon."			
RL	Plant Sci. 161:165-172(2001).			
DR	EMBL; AJ251872; CAB88871.1; -			
DR	InterPro: IPR002182; NB-ARC.			
DR	Pfam: PF00931; NB-ARC; 1.			
FT	NON_TER 1			
FT	NON_TER 1			
FT	NON_TER 1			
SO	SEQUENCE	91 AA;	10139 MW;	FD663CB318C1B881 CRC64;
	Query Match	37.4%;	Score 40;	DB 10; Length 91;
	Best Local Similarity	33.3%;	Pred. No. 39;	
	Matches	6; Conservative	6; Mismatches	6; Indels 0; Gaps 0;
QY	4	EQVEVPGSOHIDSOKRAI 21		
		: : : : : : :		
Db	64	FKMSCPSSDYIDLSKRAV 81		
		: : : : : : :		
	RESULT 5			
	Q36044			
ID	Q36044	PRELIMINARY;	PRT;	45 AA.
AC	Q36044;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	UBIQUITIN (FRAGMENT).			
OS	Mastigamoeba balaamuthi.			
OC	Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.			
OX	NCBI_TaxID=108607;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Westmann C., Hackstein J., Bakker-Grunwald T.;			
RA	"Ubiquitin sequences and polyubiquitin genes in Phreatamoeba balaamuthi			
RT	and Psalteriomonas lanterna: Phylogenetic implications."			
RL	Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AJ000658; CAA04206.1; -			
DR	HSSP; P02248; IUBI.			
DR	InterPro: IPR000626; UBIQUITIN.			
DR	Pfam: PF00240; ubiquitin; 1.			
DR	PRINTS; PR00348; UBIQUITIN.			
DR	SMART; SM00213; UBO; 1.			
DR	PROSITE; PS00299; UBIQUITIN_1; 1.			
DR	PROSITE; PS00553; UBIQUITIN_2; 1.			



us-09-786-648-5.lenlim.rspt

```
DR PRINTS; PRO0348; UBIQUITIN.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
```

KW	Nuclear protein.	1
FT	NON_TER	1
FT	NON_TER	62

Query Match	36.48;
Best Local Similarity	42.98;

```
QY      1 GETFQVEVPGSQHIDSQKAI 21
        | : | : | | | : | : |
Db      3 GKTITLVEVPSDSIDNVKQKI 23
```

RESULT 8  
036043

AC	036043;
DT	01-JAN-1998 (TREMblrel, 05, Created)
DT	01-JAN-1998 (TREMblrel, 05, last sequence update)
DT	01-Dec-2001 (TREMblrel, 10, last annotation update)

DE OBIDIOTIN (FRAGMENT).  
OS Mastigamoeba balamuthi.  
OC Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba  
CY NCBI TaxID-108607.

RN [1]  
RP SEQUENCE FROM N.A.  
RA Westmann C., Hackstein J., Bakker-Grunwald T.;  
RT "Ubiquitin sequences and polyubiquitin genes in phreatomys balamuthi"

RT and psalterionomas lanterna:  
RL Submitted (JUL-1997) to the E  
DR EMBL; AJ000657; CAA04205.1; -  
DR HSSP: P02248; 1UPI.

DR Pfam; PF00240; ubiquitin; 1.  
DR PRINTS; PR00348; UBIQUITIN.  
DR SMART; SM00213; UBO; 1.

DR	POSITE; PS00255; UBIQUITIN_	
DR	POSITE; PS50053; UBIQUITIN_2	
FT	NON_TER	1
FT	NON_TER	62

SEQUENCE	02 AM, 0000 PM,
Query Match	36.48;
Best Local Similarity	42.98;

QY 1 GETFOVEVPGSQHIDSQKAI 21  
1:1 :11 1 11: 1: 1

## RESULT 9

ID	Q9QYJ8	PRELIMINARY;
AC	Q9QYJ8;	
DT	01-MAY-2000	(TREMBLrel. 13, C

DT 01-DEC-2001 (TREMBlrel. 19, L  
DE DECAY ACCELERATING FACTOR (FF  
OS Rattus norvegicus (Rat).  
nubius) (Veternary Observations)

```

OC Mammalia; Eutheria; Rodentia;
OX NCBI_TaxID=10116;
RN [1]

```

RC STRAIN=WISTARST;  
RX MEDLINE=20130146; Pubmed=1066  
RA Miwa T., Okada N., Okada H.;

DR PRINTS, PR00348; UBIQUITIN.  
DR SMART; SM00213; UBO; 1.  
DR PROSITE; PS00239; UBIQUITIN\_1; 1  
DR PROSITE; PS50053; UBIQUITIN\_2; 1  
KW Nuclear protein.

SQ SEQUENCE 62 AA; 6978 MW; 7C14ED7613B19D20 CRC64;

Query Match	36.4%	Score 39;	DB 5;	Length 62;
Best Local Similarity	42.9%	Pred. No. 37;		
Matches 9, Conservative	4;	Mismatches 8;	Indels 0;	Gaps 0;

```
QY 1 GETFOVEVPGSQHIDSQKAI 21
    | : | : | | | : | : |
Db 3 GKTITLEVEPSPDSIDNVKQKI 23
```

RESULT 8	
036043	
ID 036043	PRELIMINARY;
AC 036043.	PRT; 62 AA

```
DT 01-JAN-1998 (TREMBlReI. 05, Created)
DT 01-JAN-1998 (TREMBlReI. 05, last sequence update)
DT 01-DEC-2001 (TREMBlReI. 19, last annotation update)
DE TR1010TTIN (FRAGMEnt)
```

OS Mastigamoeba balamuthi.  
OC Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba  
OX NCBI\_TaxID=108607;  
RN [1]

RT and *Psalteriomonas lanterna*: Phylogenetic implications.";  
RT "Ubiquitin sequences and polyubiquitin genes in *Phreatamoeba balamuthi*  
RP Westmann C., Hackstein J., Bakker-Grünwald T.;"  
RA "Ubiquitin sequences and polyubiquitin genes in *Phreatamoeba balamuthi*  
RP Westmann C., Hackstein J., Bakker-Grünwald T.;"  
RP SEQUENCE FROM N.A.

```

DR EMBL; AJ000657; CAA04205.1; -
DR HSSP; P02248; 1UBI.
DR InterPro; IPR000626; Ubiquitin

```

```
DR      Iam, FE00240, ubiquitin, 1.  
DR      PRINTS; PR00348; UBIQUITIN.  
DR      SMART; SM00213; UBQ; 1.  
DR      PROSITE; PS00299; UBIQUITIN_1
```

FT	NON_TER	1	1
FT	NON_TER	62	62
SQ	SEQUENCE	62 AA;	6906 MN;

Query Match	36.4%	Score 39;	DB 5;	Length 62;
Best Local Similarity	42.98%	Pred. No. 37;		
Matches 9; Conservative	4;	Mismatches	8;	Indels 0;
				Gaps 0;

```
QY      1 GETFQVEVPGSQHDISKAI 21
          1:1 :11 1 11: 1: 1
Db      3 GKTITLEVEPDSIDNVKQKI 23
```

RESULT 9	
Q9QYJ8	
ID Q9QYJ8	PRELIMINARY;
2	PRT; 77 AA

DT	01-MAY-2000 (Tremblurel. 13, Created)
DT	01-MAY-2000 (Tremblurel. 13, last sequence update)
DT	01-DEC-2001 (Tremblurel. 19, last annotation update)

05 Rattus norvegicus (Rat).  
0C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
0C Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=W1STARST;
BY      MEDLINE=20130146. PubMed=10663575.

```

RA Miwa T., Okada N., Okada H.;

```

RT "Alternative exon usage in the 3' region of a single gene generates
RT glycosylphosphatidylinositol-anchored and transmembrane forms of rat
RT decay-accelerating factor."
RL Immunogenetics 51:129-137(2000).
DR EMBL; AB026905; BAA88994.1; -.
FT NON_TER
FT NON_TER
SQ SEQUENCE 77 AA: 8409 MW; B822DC894E684593 CRC64;

Query Match
Best Local Similarity 36.4%; Score 39; DB 11; Length 77;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 5 QVEPVGSOHIDSOKKAI 21
Db 50 EVKVPATQHPVSKTV 66

RESULT 10
Q9PI64 PRELIMINARY; PRT; 77 AA.
AC Q9PI64;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ACTL CARRIER PROTEIN (ACP).
GN ACP OR CU0441.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
NCBI_TaxID=197;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=NCCTC 11168;
RX MEDLINE=20150912; Pubmed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jørgels K., Kariyasev A.V., Moule S., Palien M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- FUNCTION: THIS PROTEIN IS THE CARRIER OF THE GROWING FATTY ACID
CC CHAIN IN FATTY ACID BIOSYNTHESIS (BT SIMILARITY).
DR EMBL; AL139075; CAB74277.1; -.
DR HSSP; P02901; IACP.
DR InterPro; IPR003231; Acyl_carrier.
DR InterPro; IPR003880; Phosphopant_attach.
DR Pfam; PF00550; pp-binding_1.
DR ProDom; PD000887; Acyl_carrier; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KM Complete proteome; Fatty acid biosynthesis; Phosphopantetheine.
SQ SEQUENCE 77 AA: 8598 MW; 772D693BDC89588F CRC64;

Query Match
Best Local Similarity 36.4%; Score 39; DB 16; Length 77;
Matches 10; Conservative 4; Mismatches 4; Indels 12; Gaps 1;

Qy 2 ETPQVEVPGS-----OHIDSOKK 19
Db 48 EKFEVEIPSDAEKLIEDVYNIIDNKK 77

RESULT 11
Q9SAT2 PRELIMINARY; PRT; 38 AA.
AC Q9SAT2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE POLYBICUITIN HOMOLOG (FRAGMENT).

```

```

OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
NCBI_TaxID=4577;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=CV. INRA 258; TISSUE=LEAF;
RX MEDLINE=96236829; Pubmed=8680303;
RA Didierjean L., Frendo P., Nasser W., Genot G., Marivet J., Burhard G.;
RT "Heavy metal responsive genes in maize: identification and comparison
RT of their expression upon various forms of abiotic stress."
RL Planta 199:1-8(1996).
DR EMBL; S82312; AAB47174.1; -.
DR HSSP; P02248; IUBI.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR PRINTS; PR00348; UBIQUITIN.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 38 AA: 4189 MW; 3C3A2A2B8CF56B5 CRC64;

Query Match
Best Local Similarity 35.5%; Score 38; DB 10; Length 38;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GETPQVEVPGSOHIDSOKKAI 21
Db 10 GKTITLEVSSDTIDNVKAKI 30

RESULT 12
Q42009 PRELIMINARY; PRT; 52 AA.
AC Q42009;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE UBIQUITIN (FRAGMENT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=FLOWER BUDS OF A. THALIANA ECOTYPE COLUMBIA C24;
RA Macho R., Quigley F., Thomas F., Yu D.Y.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z18425; CAA79186.1; -.
DR HSSP; P02248; IUBI.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR PRINTS; PR00348; UBIQUITIN.
DR SMART; SM00213; UBO; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE 52 AA: 5914 MW; B7E64CC5347B1C44 CRC64;

Query Match
Best Local Similarity 35.5%; Score 38; DB 10; Length 52;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GETPQVEVPGSOHIDSOKKAI 21
Db 5 GKTITLEVSSDTIDNVKAKI 25

RESULT 13
Q42001 PRELIMINARY; PRT; 54 AA.
AC Q42001;
DT 042001

```

DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE UBIQUITIN (FRAGMENT).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FLOWER BUDS OF A. THALIANA ECOTYPE COLUMBIA C24;  
 RA Maché R., Quigley F., Thomas F., Yu D.Y.;  
 RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z18379; CAA79175.1; -.  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 2.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 FT NON\_TER 1 1  
 FT SEQUENCE 54 AA; 5918 MW; 35FDD6595BF63D5 CRC64;  
 SQ  
 Query Match 35.5%; Score 38; DB 10; Length 54;  
 Best Local Similarity 42.9%; Pred. No. 46;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 Db 21 GKTTLEVESDPTIDNVKAKI 41

## RESULT 14

ID Q9M646 PRELIMINARY; PRT; 72 AA.  
 AC Q9M646;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE UBIQUITIN RIP-20 (FRAGMENT).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANICLE;  
 RA Sheoran I.S., Saxena D., Saini H.S.;  
 RT "Expression of drought-responsive ubiquitin genes in reproductive  
 tissues of rice."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF216530; AAF37225.1; -.  
 DR HSSP; P02248; IUBI.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR PRINTS; PR00348; UBIQUITIN.  
 DR SMART; SM00213; UBQ; 1.  
 DR PROSITE; PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 FT NON\_TER 1 1  
 FT SEQUENCE 72 AA; 8111 MW; DCBCEDA1E0EDF85 CRC64;  
 SQ

Query Match 35.5%; Score 38; DB 10; Length 72;  
 Best Local Similarity 42.9%; Pred. No. 64;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 Db 5 GKTTLEVESDPTIDNVKAKI 25

## RESULT 15

O941M2  
 ID O941M2 PRELIMINARY; PRT; 76 AA.  
 AC O941M2;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE UBIQUITIN (FRAGMENT).  
 OS Medicago truncatula (Barrel medic).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales; Medicago.  
 OX NCBI\_TaxID=3880;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20331592; PubMed=10875337;  
 RA Salzer P., Bonanomi A., Beyer K., Voegelé-Lange R., Aeschbacher R.A.,  
 RA Lange J., Wismen A., Kim D., Cook D.R., Boller T.;  
 RT "Differential expression of eight chitinase genes in Medicago  
 truncatula roots during mycorrhiza formation, nodulation and pathogen-  
 infection."  
 RL MOL. Plant Microbe Interact. 13:763-777(2000).  
 DR EMBL; AJ24511; CAB96875.1; -.  
 FT NON\_TER 76 76  
 FT SEQUENCE 76 AA; 8525 MW; 852936277CD8403E CRC64;  
 SQ  
 Query Match 35.5%; Score 38; DB 10; Length 76;  
 Best Local Similarity 42.9%; Pred. No. 69;  
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSOHIDSOKKAI 21  
 Db 10 GKTTLEVESDPTIDNVKAKI 30

Search completed: October 24, 2002, 15:36:53  
 Job time : 25.4426 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:17 : Search time 33.7377 Seconds  
(without alignments)  
69.138 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107  
Sequence: 1 GETFQVEVPGSHIDSOKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	21	AAV87463	E. coli heat labil
2	107	100.0	93	AAV72545	ADP-ribosylating t
3	107	100.0	93	AAV41816	Escherichia coli v
4	107	100.0	93	AAW95226	E. coli heat-labli
5	107	100.0	93	AAV68365	Heat labile toxin
6	107	100.0	93	AAV66239	E coli verotoxin-1
7	107	100.0	103	AAV62379	E. coli LTB protei
8	107	100.0	123	AAV62377	E. coli LTB protei
9	107	100.0	124	AAV62373	E. coli LTB protei
10	107	100.0	124	AAV62375	E. coli LTB protei
11	107	100.0	124	AAV62380	E. coli LTB protei

12	107	100.0	134	AAV73241	Recombinant exotox
13	107	100.0	142	AAV73242	Recombinant exotox
14	107	100.0	155	AAV73243	Recombinant exotox
15	107	100.0	163	AAV73244	Recombinant exotox
16	107	100.0	371	AAV01300	Labile toxin (LT-B
17	107	100.0	371	AAV01300	C. jejuni flagellin
18	107	100.0	371	AAV01300	Cholera toxin B/en
19	107	94.4	21	AAV87462	Synthetic cholera
20	107	94.4	103	AAV04857	Heat labile entero
21	107	94.4	103	AAV94939	Heat labile entero
22	107	94.4	103	AAV06606	Cholera toxin B su
23	107	94.4	103	AAV06607	Cholera toxin B su
24	107	94.4	103	AAV08088	Amino acid sequenc
25	107	94.4	103	AAV62369	V. cholera cholera
26	107	94.4	105	AAV62369	V. cholera cholera
27	107	94.4	118	AAV04163	Cholera toxin B-su
28	107	94.4	123	AAV62374	E. coli LTB protei
29	107	94.4	124	AAV93561	B subunit of the h
30	107	94.4	124	AAV06605	Cholera toxin B su
31	107	94.4	124	AAV59770	Amino acid sequenc
32	107	94.4	124	AAV96652	Plant-optimized E.
33	107	94.4	124	AAV65992	Cholera toxin B su
34	107	94.4	124	AAV62359	V. cholera strain
35	107	94.4	124	AAV62363	V. cholera cholera
36	107	94.4	124	AAV62368	V. cholera cholera
37	107	94.4	124	AAV62376	E. coli LTB protei
38	107	94.4	124	AAV62378	E. coli LTB protei
39	107	94.4	125	AAV62372	E. coli LTB protei
40	107	94.4	131	AAV04825	Sequence of LT-B-M
41	107	94.4	138	AAV50227	Sequence of LT-B-M
42	107	94.4	170	AAV94082	LTB-CTP fusion pro
43	107	94.4	380	AAV00507	E. coli heat-labli
44	107	94.4	382	AAV00506	E. coli heat-labli
45	107	94.4	405	AAV11272	HSV-1 antigen/heat

## ALIGNMENTS

RESULT 1  
AAV87463  
ID AAV87463 standard; peptide; 21 AA.  
AC AAV87463:  
DT 03-JUL-2000 (first entry)  
XX  
DE E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.  
KW Heat labile enterotoxin subunit B; EtXB;  
KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
KW adjuvant; immune disorder; diarrhoea.  
XX  
OS Escherichia coli.  
XX  
XX  
XX W0200014114-A1.  
XX  
XX 16-MAR-2000.  
XX  
XX 07-SEP-1999; 99WO-G802970.  
XX  
XX 07-SEP-1998; 98GB-0019484.  
XX  
XX (UYBR-) UNIV BRISTOL.  
XX  
XX Williams NA, Hirst TR,  
XX  
XX WPI, 2000-256943/22.  
XX  
XX  
XX Derivatives of Escherichia coli heat labile enterotoxins useful as  
XX immunomodulators and for treating diarrhoea and which do not bind the  
XX glycolipid receptor GM-1 -

BS Disclosure: Page 15; 62pp; English.  
 XX  
 CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (CtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.  
 CC  
 SQ Sequence 21 AA:  
 OY 1 GETFOVEVPGSHIDSOKKAI 21 100.0%; Score 107; DB 21; Length 21;  
 Db 1 GETFOVEVPGSHIDSOKKAI 21 100.0%; Pred. No. 2.6e-11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 2  
 AAR72545  
 ID AAR72545 standard; peptide: 93 AA.  
 AC  
 XX AAR72545;  
 DT 28-NOV-1995 (first entry)  
 DE ADP-ribosylating toxin (verotoxin-1 B-subunit).  
 XX  
 KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;  
 KW active site; E. coli heat labile toxin; verotoxin-1;  
 KW Bordetella pertussis vaccines.  
 XX  
 OS Bacteria sp.  
 XX  
 PN EP646599-A.  
 PD 05-APR-1995.  
 XX  
 PF 23-AUG-1994; 94EP-0306219.  
 XX  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PA (UYAL-) UNIV ALBERTA.  
 XX  
 PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 PI Oomen R, Reed RJ, Stein PE;  
 XX  
 DR WPI: 1995-132623/18.  
 XX  
 PT New modified forms of pertussis holotoxin - developed using  
 PT crystalline forms of pertussis holotoxin and its complexes with  
 PT other molecules  
 XX  
 SS Disclosure: Fig 5; 54pp; English.

```

XX AAR72540-R72545 are structurally equivalent B-subunits from three
CC ADP-ribosylating toxins; pertussis holotoxin (PT), E. coli heat
CC labile toxin (LT), and verotoxin-1 (VT). The structural
CC information obtd. from these comparisons was used to identify
CC sites which contribute to PT's biological activity. By modifying
CC these sites the claimed pt mutants of the invention were produced,
CC they can be used in the development of vaccines against Bordetella
CC pertussis infection.
CC XX
SQ Sequence 93 AA:

Query Match 100.0%; Score 107; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 1,5e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSOHDSOKKAI 21
   |||||
Db 35 GETFOVEVPGSHIDSOKKAI 55

RESULT 3
AAV41816
ID AAV41816 standard; peptide: 93 AA.
XX
XX AAV41816;
AC
XX
DT 08-DEC-1999 (first entry)
XX
DE Escherichia coli verotoxin-1 B-subunit.
XX
XX ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;
KM three-dimensional structure; LT; immunoprotective; Infection.
XX
OS Escherichia coli.
XX
PN US5965385-A.
XX
PD 12-OCT-1999.
XX
PE 06-JUN-1995; 95US-0467974.
XX
PR 22-AUG-1994; 94US-0292968.
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (UTAL-) UNIV ALBERTA.
PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
PI Hazes B, Oomen RP;
DR WPI: 1999-579908/49.
XX
XX New method for producing modified pertussis holotoxin -
XX
XX Example 3: Fig 5; 4pp: English.
XX
XX A method has been developed of producing a modified pertussis holotoxin,
XX involving analysis of the 3-dimensional form of the crystalline
XX holotoxin. The pertussis holotoxin modification process comprises:
XX (1) identification of at least one amino acid (aa) residue of the
XX holotoxin for modification by analysing the 3-dimensional form of the
XX crystalline holotoxin, in relation to known information of the protein
XX structure and function; (2) effecting mutagenesis (by removing or
XX replacing a nucleotide sequence encoding at least one (aa)) of a tox
XX operon; and (3) expressing mutant tox box in a Bordetella organism to
XX produce the modified holotoxin. This method is used for modifying
XX pertussis holotoxin, by studying its 3-dimensional crystalline
XX structure. Modifying the holotoxin, alters its biological properties.
XX By analysing the 3-dimensional crystalline structure of the pertussis
XX holotoxin, functional (aa) which affect biological properties of the
XX pertussis holotoxin can be identified. This can be used to predict (aa)

```

CC which contribute to the toxicity of the holotoxin to produce  
 CC immunoprotective, genetically-detoxified analogues of pertussis  
 CC holotoxin. The present sequence represents an ADP-ribosylating toxin  
 CC B-subunit peptide used in the exemplification of the present  
 CC invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 107; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHIDSOKKAI 21  
 |||||  
 Db 35 GEFQVEVPGSOHIDSOKKAI 55

#### RESULT 4

AAW95226  
 ID AAW95226 standard; peptide; 93 AA.

AC AAW95226;

DT 16-MAR-1999 (first entry)

DE E. coli heat-labile toxin (LT) beta-subunit sequence.

KM pertussis holotoxin; PT; modified; effector; toxicity; cell binding;  
 KM enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;  
 KM structural analysis; interacting site; mitogenicity; adjuvanticity;  
 KM heat-labile; LT.

OS Escherichia coli.

PN US5856122-A.

PD 05-JAN-1999.

PF 22-AUG-1994; 94US-0292968.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (UNIV-) UNIV ALBERTA.

PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;  
 PI Oomen RP, Read RJ, Stein PE;

DR WPI: 1999-105104/09.

PT Modifications to e.g. enzymatic activity, mitogenicity and cell  
 PT binding of pertussis holotoxin - by identifying interaction sites of  
 PT a molecule with crystalline toxin and modifying the identified site  
 PS Example 3; Fig 5; 40pp; English.

CC The invention relates to methods of preparing a pertussis holotoxin (PT)  
 CC having a modified biological activity. One method comprises identifying  
 CC at least 1 site in a PT that interacts with a molecule that is capable of  
 CC forming a complex with the holotoxin and which molecule is an effector  
 CC molecule which is an adenine nucleotide and which site contributes to  
 CC toxicity, cell binding or enzymatic activity of PT. The functional  
 CC interacting site(s) are identified by analysing the three dimensional  
 CC structure of crystalline PT, determined by X-ray crystallography. The  
 CC identified interacting site(s) are modified to alter toxicity, cell  
 CC binding or enzyme activity of the PT. The methods can be used to alter a  
 CC biological activity such as toxicity, enzymatic activity, mitogenicity,  
 CC cell binding and adjuvanticity of the PT. The three-dimensional structure  
 CC of PT have functional and/or structural resemblance to other bacterial  
 CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the  
 CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present  
 CC sequence represents the beta-subunit of LT toxin.

XX  
 SQ Sequence 93 AA;

Query Match 100.0%; Score 107; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEFQVEVPGSOHIDSOKKAI 21  
 |||||  
 Db 35 GEFQVEVPGSOHIDSOKKAI 55

#### RESULT 5

AA68365  
 ID AAY68365 standard; Peptide; 93 AA.

AC AAY68365;

DT 17-APR-2000 (first entry)

DE Heat labile toxin B subunit SEQ ID NO:26.

KM pertussis toxin; heat labile toxin; verotoxin; exotoxin A; holotoxin;  
 KM diphtheria toxin; ADP-ribosylating toxin; mannose binding protein;  
 KM infection; crystal structure; X-ray crystallography; detoxification;  
 KM immunogenic.

OS Escherichia coli.

PN US6018022-A.

PD 25-JAN-2000.

PF 06-JUN-1995; 95US-0467976.

PR 22-AUG-1994; 94US-0292968.

PR 24-AUG-1993; 93US-0110947.

PR 31-MAY-1994; 94US-0251121.

PA (CONN-) CONNAUGHT LAB LTD.

PA (UNIV-) UNIV ALBERTA.

PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;  
 PI Hazes B, Oomen RP;

DR WPI: 2000-136703/12.

PT Crystalline form of isolated pertussis holotoxin useful in studying  
 PT proteins which have functional resemblance -  
 PS Example 3; Fig 5; 42pp; English.

CC The present invention describes a crystalline form of isolated  
 CC pertussis holotoxin, in which the molecules of pertussis toxin have  
 CC a three dimensional structure represented in the specification,  
 CC complexed with a polysaccharide molecule capable of forming a complex  
 CC with the holotoxin. The crystalline form of the pertussis holotoxin  
 CC can be used in a comparison with other proteins which have functional  
 CC resemblance to pertussis holotoxin with the aim of modifying other  
 CC proteins. Identifying the unknown sites of toxicity by comparison  
 CC with the three dimensional structure of pertussis holotoxin provides a  
 CC technique for detoxification of toxins to produce useful immunogenic  
 CC but non-toxic analogues. It can also be used as a primary standard to  
 CC measure the quantity, purity or efficacy of less pure compositions  
 CC containing pertussis toxin. AAY68340 to AAY68385 represent peptides  
 CC used in the exemplification of the present invention.

SQ Sequence 93 AA;

Query Match 100.0%; Score 107; DB 21; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEYPSQIHDSOKKAI 21  
 ||||||||||||||||||  
 DB 35 GETFOVEYPSQIHDSOKKAI 55

RESULT 6:  
 AAB62379  
 ID AAB66239 standard; Protein: 93 AA.  
 XX  
 AC AAB66239;  
 XX  
 DT 03-APR-2001 (first entry)  
 XX  
 DE E coli verotoxin-1 B subunit SEQ ID NO: 26.  
 XX  
 KM Pertussis toxin; crystal structure; whooping cough; biological activity;  
 KM lymphocytosis-promoting factor; histamine-sensitising factor;  
 KM islet-activating protein.  
 XX  
 OS Escherichia coli.  
 XX  
 PN US6168928-B1.  
 XX  
 PD 02-JAN-2001.  
 XX  
 PE 21-MAY-1998; 98US-0082514.  
 XX  
 PR 22-AUG-1994; 94US-0292968.  
 PR 24-AUG-1993; 93US-0110947.  
 PR 31-MAY-1994; 94US-0251121.  
 XX  
 PA (CONN-) CONNACHT LAB LTD.  
 PI Read RJ, Cockle SA, Oomen RP, Loosmore S, Klein MH, Armstrong GD;  
 PI Haze B, Stein PE;  
 XX  
 DR WPI: 2001-122260/13.  
 XX  
 PT Modifying pertussis holotoxin to produce detoxified PT analogs,  
 PT comprising analyzing crystalline structure of toxin, to identify sites  
 PT of toxicity, cell binding or enzyme activity of PT and modifying  
 PT identified site -  
 XX  
 PS Example 3; Fig 5; 41pp; English.  
 XX  
 CC The present invention provides a method for producing a pertussis toxin  
 CC (also designated lymphocytosis-promoting factor, histamine-sensitising  
 CC factor and islet activating protein) with a modified biological activity,  
 CC involving analysing the crystal structure of the protein to identify  
 CC active sites which can then be modified. This may lead to an alteration  
 CC in the toxicity, cell binding or enzyme activity of the toxin. This can  
 CC be used in the production of immunoprotective analogues of pertussis  
 CC toxin. Pertussis toxin is the cause of whooping cough following infection  
 CC by Bordetella pertussis.  
 XX  
 SQ Sequence 93 AA:  
 XX

Query Match 100.0%; Score 107; DB 22; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEYPSQIHDSOKKAI 21  
 ||||||||||||||||||  
 DB 35 GETFOVEYPSQIHDSOKKAI 55

RESULT 7  
 AAB62379  
 ID AAB62379 standard; Protein: 103 AA.  
 XX  
 AC AAB62379;  
 XX  
 DT 29-JUN-2001 (first entry)  
 XX

XX  
 DE E. coli LTB protein variant (GI: 494265).  
 XX  
 KM ABB subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 KM immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;  
 KM infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;  
 KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200127144-A2.  
 XX  
 PD 19-APR-2001.  
 XX  
 PE 05-OCT-2000; 2000WO-US27607.  
 XX  
 PR 08-OCT-1999; 99US-0158561.  
 XX  
 PA (ACT-) ACTIVE BIOTECH AB.  
 XX  
 PI Handley HH, Haaparanta T, Ewalt KL;  
 XX  
 DR WPI: 2001-281974/29.  
 XX  
 PE Recombinant ABB subunit protein comprising a mutation that alters the  
 PT number residues available for chemical modification, useful for  
 PT covalently linking to an immunogen or vaccine which can be used for  
 PT treating autoimmune diseases -  
 XX  
 PS Disclosure; Page 72-73; 78pp; English.  
 XX  
 CC The invention relates to a recombinant ABB subunit protein comprising at  
 CC least one mutation, where the mutation alters the number of amino acid  
 CC residues available for chemical modification as compared to a wild type  
 CC ABB subunit protein, and where the recombinant protein retains an  
 CC effective target ligand binding affinity. A recombinant ABB subunit  
 CC protein such as cholera toxin B protein (CTB) can be specifically  
 CC covalently linked at lysines or cysteines to an immunogen or vaccine.  
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron  
 CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the  
 CC invention can also be used to induce tolerance to infection, e.g.  
 CC parasitic infection. The present sequence represents a variant of the  
 CC E. coli heat labile toxin B (LTB) protein.  
 XX  
 SQ Sequence 103 AA;  
 XX

Query Match 100.0%; Score 107; DB 22; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEYPSQIHDSOKKAI 21  
 ||||||||||||||||||  
 DB 45 GETFOVEYPSQIHDSOKKAI 65

RESULT 8  
 AAB62377  
 ID AAB62377 standard; Protein: 123 AA.  
 XX  
 AC AAB62377;  
 XX  
 DT 29-JUN-2001 (first entry)  
 XX  
 DE E. coli LTB protein variant (GI: 223254).  
 XX  
 KM ABB subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 KM immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;  
 KM infection; parasitic; immunosuppressive; antiarthritic; antirheumatic;  
 KM antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.  
 XX  
 OS Escherichia coli.  
 XX



PN WO200127144-A2.  
 XX  
 XX 19-APR-2001.  
 XX  
 XX 05-OCT-2000; 2000WO-US27607.  
 XX  
 XX 08-OCT-1999; 99US-0158561.  
 XX  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 XX Handley HH, Haaparanta T, Ewalt KL;  
 XX  
 XX WPI: 2001-281974/29.  
 XX  
 XX  
 XX Recombinant AB5B subunit protein comprising a mutation that alters the  
 PT number residues available for chemical modification, useful for  
 PT covalently linking to an immunogen or vaccine which can be used for  
 PT treating autoimmune diseases -  
 XX  
 XX  
 XX Disclosure; Page 72; 78pp; English.  
 XX  
 XX The invention relates to a recombinant AB5B subunit protein comprising at  
 CC least one mutation, where the mutation alters the number of amino acid  
 CC residues available for chemical modification as compared to a wild type  
 CC AB5B subunit protein, and where the recombinant protein retains an  
 CC effective target ligand binding affinity. A recombinant AB5B subunit  
 CC protein such as cholera toxin B protein (CTB) can be specifically  
 CC covalently linked at lysines or cysteines to an immunogen or vaccine.  
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron  
 CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the  
 CC invention can also be used to induce tolerance to infection, e.g. a  
 CC parasitic infection. The present sequence represents a variant of the  
 CC E. coli heat labile toxin B (LTB) protein.  
 XX  
 XX  
 XX Sequence 123 AA;  
 SQ  
 Query Match 100.0%; Score 107; DB 22; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSQHDSQKRAI 21  
 ||||||||||||||||  
 Db 65 GETFOVEVPGSQHDSQKRAI 85  
 RESULT 9  
 AAB62373  
 ID AAB62373 standard; Protein; 124 AA.  
 XX  
 XX AAB62373;  
 XX  
 XX 29-JUN-2001 (first entry)  
 XX  
 XX E. coli LTB protein variant (GI: 1169505).  
 DE  
 XX  
 XX AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;  
 KW infection; parasitic; immunosuppressive; antirheumatic; antineumatic;  
 KW antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.  
 XX  
 XX Escherichia coli.  
 OS  
 XX WO200127144-A2.  
 PN  
 XX  
 XX 19-APR-2001.  
 PD  
 XX  
 XX 05-OCT-2000; 2000WO-US27607.  
 PF  
 XX 08-OCT-1999; 99US-0158561.  
 PR  
 XX  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 PA  
 XX

PI Handley HH, Haaparanta T, Ewalt KL;  
 XX  
 XX WPI: 2001-281974/29.  
 XX  
 XX  
 XX Recombinant AB5B subunit protein comprising a mutation that alters the  
 PT number residues available for chemical modification, useful for  
 PT covalently linking to an immunogen or vaccine which can be used for  
 PT treating autoimmune diseases -  
 XX  
 XX  
 XX Disclosure; Page 70; 78pp; English.  
 XX  
 XX The invention relates to a recombinant AB5B subunit protein comprising at  
 CC least one mutation, where the mutation alters the number of amino acid  
 CC residues available for chemical modification as compared to a wild type  
 CC AB5B subunit protein, and where the recombinant protein retains an  
 CC effective target ligand binding affinity. A recombinant AB5B subunit  
 CC protein such as cholera toxin B protein (CTB) can be specifically  
 CC covalently linked at lysines or cysteines to an immunogen or vaccine.  
 CC Recombinant mutant CTB (rCTB) can be used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid arthritis, encephalomyelitis (or other neuron  
 CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the  
 CC invention can also be used to induce tolerance to infection, e.g.  
 CC parasitic infection. The present sequence represents a variant of the  
 CC E. coli heat labile toxin B (LTB) protein.  
 XX  
 XX  
 XX Sequence 124 AA;  
 SQ  
 Query Match 100.0%; Score 107; DB 22; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GETFOVEVPGSQHDSQKRAI 21  
 ||||||||||||||||  
 Db 66 GETFOVEVPGSQHDSQKRAI 86  
 RESULT 10  
 AAB62375  
 ID AAB62375 standard; Protein; 124 AA.  
 XX  
 XX AAB62375;  
 XX  
 XX 29-JUN-2001 (first entry)  
 XX  
 XX E. coli LTB protein variant (GI: 145833).  
 DE  
 XX  
 XX AB5B subunit protein; mutation; recombinant; cholera toxin B; CTB;  
 KW immunogen; autoimmune disease; neuron demyelinating disease; rCTB; LTB;  
 KW infection; parasitic; immunosuppressive; antirheumatic; antineumatic;  
 KW antidiabetic; neuroprotective; vaccine; heat labile toxin B; variant.  
 XX  
 XX Escherichia coli.  
 OS  
 XX WO200127144-A2.  
 PN  
 XX  
 XX 19-APR-2001.  
 PD  
 XX  
 XX 05-OCT-2000; 2000WO-US27607.  
 PF  
 XX 08-OCT-1999; 99US-0158561.  
 PR  
 XX  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 PA  
 XX Handley HH, Haaparanta T, Ewalt KL;  
 XX  
 XX WPI: 2001-281974/29.  
 DR  
 XX  
 XX Recombinant AB5B subunit protein comprising a mutation that alters the  
 PT number residues available for chemical modification, useful for  
 PT covalently linking to an immunogen or vaccine which can be used for  
 PT treating autoimmune diseases -  
 XX  
 XX  
 XX Disclosure; Page 71; 78pp; English.  
 PS

```

CC demyelinating diseases) and diabetes. The rCTB or other B subunits of the
CC invention can also be used to induce tolerance to infection, e.g. the
CC parasitic infection. The present sequence represents a variant of the
CC E. coli heat labile toxin B (LTB) protein.
CC
CC
CC
CC
SQ Sequence 124 AA;

Query Match 100.0%; Score 107; DB 22; Length 124;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHIDSOKKAI 21
   |||||||||||||||||||
Db 66 GETFOVEVPGSHIDSOKKAI 86

RESULT 12
AAB73241
ID AAB73241 standard; Protein; 134 AA.
XX
XX AAB73241;
XX
XX DT 14-MAY-2001 (first entry)
XX
XX DE Recombinant exotoxin protein variant LTbPL.
XX
XX Exotoxin mucosal cell binding motif; nucleic acid delivery;
XX nucleic acid affinity domain; heat-labile enterotoxin.
XX
XX OS Unidentified.
XX
XX PN WO20011960-A1.
XX
XX PD 22-FEB-2001.
XX
XX PE 18-AUG-2000; 2000WO-US22715.
XX
XX PR 18-AUG-1999; 99US-0149294.
XX
XX PA (AGRI-) AGRIVAX INC.
XX
XX Welter LM;
XX
XX PI WPI: 2001-211103/21.
XX
XX DR N-PSDB; AAF75712.
XX
XX
XX Novel exotoxin protein variant useful as protein carrier for
XX facilitating gene delivery, comprises a mucosal cell binding motif of
XX an exotoxin and a nucleic acid affinity domain
XX
XX Example 1; Fig 9; 57Pp; English.
XX
XX The present invention relates to recombinant exotoxin protein variants,
XX which comprise an exotoxin mucosal cell binding motif and a nucleic acid
XX affinity domain. The present sequence is one such protein variant. In the
XX present invention the heat-labile enterotoxin (LT) of Escherichia coli
XX was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin
XX (bARE) family. The protein variants are useful for selectively delivering
XX nucleic acid to mucosal cells, for inducing an immune response when the
XX nucleic acid encodes an antigen to which the immune response is desired,
XX for selectively delivering a gene to a mucosal cell, and for achieving
XX expression of a protein in a subject, by administering a composition
XX comprising the protein variant.
XX
XX
XX Sequence 134 AA;

Query Match 100.0%; Score 107; DB 22; Length 134;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHIDSOKKAI 21
   |||||||||||||||||||
Db 66 GETFOVEVPGSHIDSOKKAI 86

```

## RESULT 13

AAB73242  
ID AAB73242 standard; Protein; 142 AA.

AC AAB73242;

DT 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTBplh.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.

XX Unidentified.

PN WO200111960-A1.

PD 22-FEB-2001.

PF 18-AUG-2000; 2000WO-US22715.

PR 18-AUG-1999; 99US-0149294.

PA (AGRI-) AGRIVAX INC.

PI Welter LM;

DR WPI; 2001-211103/21.

DR N-PSDB; AAF75713.

PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain

PS Example 1; Fig 10; 57pp; English.

CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of *Escherichia coli*  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bARE) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.

XX Sequence 142 AA;

Query Match 100.0%; Score 107; DB 22; Length 142;  
Best Local Similarity 100.0%; Pred. No. 2.5e-10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEFQVEVPSQHIDSQKAI 21

DB 66 GEFQVEVPSQHIDSQKAI 86

## RESULT 14

AAB73243  
ID AAB73243 standard; Protein; 155 AA.

AC AAB73243;

DT 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTB-P.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin.

XX Unidentified.

PN WO200111960-A1.

PD 22-FEB-2001.

PF 18-AUG-2000; 2000WO-US22715.

PR 18-AUG-1999; 99US-0149294.

PA (AGRI-) AGRIVAX INC.

PI Welter LM;

DR WPI; 2001-211103/21.

DR N-PSDB; AAF75714.

PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain

PS Example 1; Fig 11; 57pp; English.

CC The present invention relates to recombinant exotoxin protein variants,  
CC which comprise an exotoxin mucosal cell binding motif and a nucleic acid  
CC affinity domain. The present sequence is one such protein variant. In the  
CC present invention the heat-labile enterotoxin (LT) of *Escherichia coli*  
CC was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin  
CC (bARE) family. The protein variants are useful for selectively delivering  
CC nucleic acid to mucosal cells, for inducing an immune response when the  
CC nucleic acid encodes an antigen to which the immune response is desired,  
CC for selectively delivering a gene to a mucosal cell, and for achieving  
CC expression of a protein in a subject, by administering a composition  
CC comprising the protein variant.

XX Sequence 155 AA;

Query Match 100.0%; Score 107; DB 22; Length 155;  
Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GEFQVEVPSQHIDSQKAI 21

DB 66 GEFQVEVPSQHIDSQKAI 86

## RESULT 15

AAB73244  
ID AAB73244 standard; Protein; 163 AA.

AC AAB73244;

DT 14-MAY-2001 (first entry)

DE Recombinant exotoxin protein variant LTB-Ph.

XX Exotoxin mucosal cell binding motif; nucleic acid delivery;  
KW nucleic acid affinity domain; heat-labile enterotoxin; ds.

XX Unidentified.

PN WO200111960-A1.

PD 22-FEB-2001.

PF 18-AUG-2000; 2000WO-US22715.

PR 18-AUG-1999; 99US-0149294.

PA (AGRI-) AGRIVAX INC.

PI Welter LM;

XX	WPI; 2001-211103/21.
DR	
DR	N-PSDB; AAF75715.
XX	

PT Novel exotoxin protein variant useful as protein carrier for  
PT facilitating gene delivery, comprises a mucosal cell binding motif of  
PT an exotoxin and a nucleic acid affinity domain -  
v

PS Example 1; Fig 12; 57pp; English.

The present invention relates to recombinant exotoxin protein variants, which comprise an exotoxin mucosal cell binding motif and a nucleic acid affinity domain. The present sequence is one such protein variant. In the present invention the heat-labile enterotoxin (LT) of *Escherichia coli* was used as an exotoxin. LT is a member of the ADP-ribosylating exotoxin (bare) family. The protein variants are useful for selectively delivering nucleic acid to mucosal cells, for inducing an immune response when the nucleic acid encodes an antigen to which the immune response is desired, for selectively delivering a gene to a mucosal cell, and for achieving expression of a protein in a subject, by administering a composition comprising the protein variant.

SQ Sequence 163 AA;

Query Match	100.0%	Score 107;	DB 22;	Length 163;
Best Local Similarity	100.0%;	Pred. NO. 3e-10;		
Matches 21; Conservative	0;	Mismatches	0;	Gaps 0;

```
QY 1 GETFQVEVPGSQHIDSQKAI 21
    |||
Db 66 GETFQVEVPGSQHIDSQKAI 86
```

```
Search completed: October 24, 2002, 15:20:28
Job time : 33.7377 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:18:53 : Search time 12.3934 Seconds  
(without alignments)  
41.388 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GEFQVEVPSQSHDSQKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCYUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	93	2	US-08-292-968-26
2	107	100.0	93	2	US-08-467-974-26
3	107	100.0	93	2	US-08-467-536-26
4	107	100.0	93	3	US-08-467-976-26
5	107	100.0	93	4	US-09-082-514-26
6	107	100.0	371	2	US-08-829-026A-6
7	101	94.4	102	3	US-08-952-337-5
8	101	94.4	102	3	US-08-952-337-6
9	101	94.4	103	2	US-08-472-171-2
10	101	94.4	103	2	US-08-894-526-2
11	101	94.4	103	2	US-09-013-047-2
12	101	94.4	103	4	US-09-374-597-2
13	101	94.4	103	4	US-09-191-852-21
14	101	94.4	103	5	PCT-US95-13376-21
15	101	94.4	123	3	US-08-952-337-1
16	101	94.4	123	3	US-08-952-337-2
17	101	94.4	124	3	US-08-747-410-2
18	89	83.2	124	1	US-08-448-045C-4
19	89	83.2	124	6	US-08-435-605A-12
20	89	83.2	124	6	5223610-3
21	54	50.5	448	2	US-08-878-989-2
22	54	50.5	448	4	US-09-272-796-2
23	54	50.5	508	4	US-09-344-700-4
24	43	40.2	1810	5	PCT-US95-11684-4
25	42	39.3	855	2	US-09-027-337-2
26	41	38.3	1022	1	US-08-271-364A-8
27	41	38.3	1022	2	US-08-223-715B-27

28	40.5	37.9	856	3	US-08-709-784-2	Sequence 2, Appl
29	40.5	37.9	862	2	US-08-209-521-23	Sequence 23, Appl
30	40.5	37.9	862	2	US-08-209-521-30	Sequence 30, Appl
31	40.5	37.9	862	4	US-09-059-461-2	Sequence 2, Appl
32	40.5	37.9	862	4	US-08-961-810-133	Sequence 133, App
33	40.5	37.9	862	4	US-08-352-902D-133	Sequence 133, App
34	40.5	37.4	774	3	US-08-902-632-2	Sequence 2, Appl
35	40.5	37.4	774	3	US-09-073-354-1	Sequence 1, Appl
36	40.5	37.4	774	3	US-08-656-005A-1	Sequence 1, Appl
37	40.5	37.4	774	4	US-09-073-258-1	Sequence 1, Appl
38	40.5	37.4	774	4	US-09-363-095-1	Sequence 1, Appl
39	40.5	37.4	774	4	US-09-418-027-1	Sequence 1, Appl
40	40.5	37.4	774	2	US-08-906-925-4	Sequence 4, Appl
41	40.5	37.4	779	1	US-08-375-134-12	Sequence 12, Appl
42	40.5	37.4	779	5	PCT-US95-15263-12	Sequence 12, Appl
43	39.5	36.9	1090	4	US-09-346-237-5	Sequence 5, Appl
44	39.5	36.4	75	1	US-08-350-884-35	Sequence 35, Appl
45	39.5	36.4	75	1	US-08-709-173-35	Sequence 35, Appl

#### ALIGNMENTS

RESULT 1  
US-08-292-968-26  
; Sequence 26, Application US/08292968  
; Patent No. 5856122  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: COOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/292,968  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-388  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21  
|||||

DB 35 GETFOVEVPGSQHIDSOKKAI 55

## RESULT 2

US-08-467-974-26

; Sequence 26, Application US/08467974

; Patent No. 5965385

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,974

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/467,536

; FILING DATE: 06-JUN-1995

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/251,121

; FILING DATE: 31-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-454 MIS:V9

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1153

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-974-26

US-08-467-974-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21  
|||||

DB 35 GETFOVEVPGSQHIDSOKKAI 55  
|||||

## RESULT 3

US-08-467-536-26

; Sequence 26, Application US/08467536

; Patent No. 5977304

; GENERAL INFORMATION:

; APPLICANT: READ, Randy J.

; APPLICANT: STEIN, Penelope E.

; APPLICANT: COCKLE, Stephen A.

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: KLEIN, Michel H.

; APPLICANT: ARMSTRONG, Glen D.

; APPLICANT: HAZES, Bart

; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/467,536

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/292,968

; FILING DATE: 22-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/110,947

; FILING DATE: 24-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I.

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-455 MIS:V9

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1153

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 93 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-467-536-26

Query Match 100.0%; Score 107; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHIDSOKKAI 21  
|||||

DB 35 GETFOVEVPGSQHIDSOKKAI 55

RESULT 4  
US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022  
; GENERAL INFORMATION:

APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,976  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-453 MIS-VG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-976-26  
Query Match 100.0%; Score 107; DB 3; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 35 GETFOVEVPGSOHIDSOKKAI 55  
RESULT 5  
US-09-082-514-26  
Sequence 26, Application US/09082514  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,514  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 24-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-810  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-082-514-26  
Query Match 100.0%; Score 107; DB 4; Length 93;  
Best Local Similarity 100.0%; Pred. No. 2.7e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSOHIDSOKKAI 21  
|||||  
DB 35 GETFOVEVPGSOHIDSOKKAI 55  
RESULT 6  
US-08-829-026A-6  
Sequence 6, Application US/0829026A  
Patent No. 5837825  
GENERAL INFORMATION:  
APPLICANT: Meinersmann, Richard J.  
APPLICANT: Khoury, Christian A.  
TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Janelle S. Graeter  
STREET: Room 411, Building 005, BARC-W  
CITY: Beltsville  
STATE: MD  
COUNTRY: USA  
ZIP: 20705  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/829,026A  
FILING DATE: 18-AUG-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Graeter, Janelle S.  
REGISTRATION NUMBER: 35,024  
REFERENCE/DOCKET NUMBER: 0106.97  
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 301-504-5676
; TELEFAX: 301-504-5060
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-829-026A-6

Query Match
Best Local Similarity 100.0%; Score 107; DB 2; Length 371;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GETFOVEVPGSQHIDSOKKAI 21
Db 49 GETFOVEVPGSQHIDSOKKAI 69

RESULT 7
US-08-952-337-5
; Sequence 5, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; EARLIER FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-08-952-337-5

Query Match
Best Local Similarity 94.4%; Score 101; DB 3; Length 102;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GETFOVEVPGSQHIDSOKKAI 21
Db 44 GATFOVEVPGSQHIDSOKKAI 64

RESULT 8
US-08-952-337-6
; Sequence 6, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; EARLIER FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 102
; TYPE: PRT
```

```

; ORGANISM: Escherichia coli
; US-08-952-337-6

Query Match
Best Local Similarity 94.4%; Score 101; DB 3; Length 102;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GETFOVEVPGSQHIDSOKKAI 21
Db 44 GATFOVEVPGSQHIDSOKKAI 64

RESULT 9
US-08-472-171-2
; Sequence 2, Application US/08472171
; Patent No. 5932714
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 330 University Avenue, Suite 701
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,171
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334
; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:v9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-171-2

Query Match
Best Local Similarity 94.4%; Score 101; DB 2; Length 103;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GETFOVEVPGSQHIDSOKKAI 21
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 10
US-08-894-526-2
; Sequence 2, Application US/08894526
; Patent No. 5942418
; GENERAL INFORMATION:
```



```

; APPLICANT: Loosmore, Sheena M
; APPLICANT: Yacoub, Reza K
; APPLICANT: Zealey, Gavin R
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,526
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-894-526-2

Query Match          94.4%; Score 101; DB 2; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHIDSOKKAI 21
Db 45 GATFOVEVPGSHIDSOKKAI 65

RESULT 11
US-09-013-047-2
; Sequence 2, Application US/09013047
; Patent No. 5998168
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,047
```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,171
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334
; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-047-2

Query Match          94.4%; Score 101; DB 2; Length 103;
Best Local Similarity 95.2%; Pred. No. 3.3e-10;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSHIDSOKKAI 21
Db 45 GATFOVEVPGSHIDSOKKAI 65

RESULT 12
US-09-374-597-2
; Sequence 2, Application US/09374597
; Patent No. 6140082
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/393,334
; FILING DATE: FEBRUARY 23, 1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1163
; TELEFAX: 065-24567 Simbas
```



Fri Oct 25 17:51:33 2002

us-09-786-648-5.ra1

Page 7

	Matches	20;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
-QY	1	GETFQVEYVPGSQHIDSQKAI	21							
	1									
Db	65	GATFQVEYVPGSQHIDSQKAI	85							

Search completed: October 24, 2002, 15:24:00  
Job time : 13.3934 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:20:33 : Search time 141.836 Seconds  
(without alignments)  
52.114 Million cell updates/sec

Title: US-09-786-648-5  
Perfect score: 107  
Sequence: 1 GETFOVEVPGSQHIDSOKKAI 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:

```
1: /cgn2_6/ptodata/1/paa/PCUS.COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097.COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098.COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099.COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100.COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101.COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US60.COMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	21	US-09-786-648-5	Sequence 5, Appli
2	107	100.0	93	US-08-110-947-10	Sequence 10, Appli
3	107	100.0	93	US-08-110-947A-26	Sequence 26, Appli
4	107	100.0	93	US-08-251-121-26	Sequence 26, Appli
5	107	100.0	371	US-08-150-305A-3	Sequence 3, Appli
6	107	100.0	371	US-08-784-218-6	Sequence 6, Appli
7	107	100.0	371	US-08-829-026-5	Sequence 5, Appli

8	101	94.4	21	US-09-786-648-4	Sequence 4, Appli
9	101	94.4	103	US-08-393-334-2	Sequence 2, Appli
10	101	94.4	103	US-08-782-832-15	Sequence 15, Appli
11	101	94.4	103	US-08-817-906-21	Sequence 21, Appli
12	101	94.4	103	US-08-836-433-14	Sequence 14, Appli
13	101	94.4	116	US-09-836-433-20	Sequence 20, Appli
14	101	94.4	119	US-09-836-433-22	Sequence 22, Appli
15	101	94.4	124	PCT-US99-30747-55	Sequence 55, Appli
16	101	94.4	124	US-09-470-124-55	Sequence 55, Appli
17	101	94.4	138	US-08-914-479-2	Sequence 2, Appli
18	101	94.4	138	US-08-914-479A-2	Sequence 2, Appli
19	101	94.4	313	US-09-756-983-15	Sequence 15, Appli
20	101	94.4	351	US-09-756-983-18	Sequence 18, Appli
21	101	94.4	364	US-09-756-983-22	Sequence 22, Appli
22	95	88.8	124	PCT-US99-30747-57	Sequence 57, Appli
23	95	88.8	124	US-09-470-124-57	Sequence 57, Appli
24	89	83.2	124	US-09-760-234-7	Sequence 7, Appli
25	89	83.2	382	PCT-US01-08582-3	Sequence 3, Appli
26	89	83.2	382	PCT-US01-08582-4	Sequence 4, Appli
27	89	83.2	461	US-09-051-315-2	Sequence 2, Appli
28	89	83.2	461	US-09-423-493-2	Sequence 2, Appli
29	89	83.2	750	US-09-402-100-2	Sequence 2, Appli
30	89	83.2	1338	US-09-402-100-4	Sequence 4, Appli
31	76	71.0	15	US-08-732-371-1	Sequence 1, Appli
32	76	71.0	15	US-08-732-371A-1	Sequence 1, Appli
33	62	57.9	12	US-09-786-648-3	Sequence 3, Appli
34	54	50.5	286	US-09-758-445-415	Sequence 415, Appli
35	54	50.5	448	US-09-769-970-2	Sequence 2, Appli
36	54	50.5	508	US-09-563-997-4	Sequence 4, Appli
37	49	45.8	248	US-09-451-320-388	Sequence 388, Appli
38	49	45.8	267	US-09-391-631-3104	Sequence 3104, Appli
39	49	45.8	267	US-09-689-980-802	Sequence 802, Appli
40	49	45.8	270	US-09-391-631-3103	Sequence 3103, Appli
41	49	45.8	270	US-09-688-980-801	Sequence 801, Appli
42	48	44.9	340	US-09-252-691-6923	Sequence 6923, Appli
43	48	44.9	484	US-09-252-691C-6923	Sequence 6923, Appli
44	48	44.9	484	US-09-252-991A-29252	Sequence 29252, Appli
45	47	43.9	369	US-09-708-427-58248	Sequence 58248, Appli

#### ALIGNMENTS

RESULT 1  
US-09-786-648-5  
Sequence 5, Application US/09786648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacc  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/GB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 5  
LENGTH: 21  
TYPE: PRT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 45...65  
OTHER INFORMATION: Isolated or synthetic EtxB beta4-alpha2 loop fragment derivabl  
US-09-786-648-5

Query Match 100.0%; Score 107; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSQHIDSOKKAI 21

Db 1 GETFOVEVPGSQHIDSOKKAI 21  
RESULT 2  
US-08-110-947-10  
; Sequence 10. Application US/08110947  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: COOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Matlare, Ltd.  
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson  
; STREET: Davis Hwy.  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/110,947  
; FILING DATE: 24-AUG-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FALLOW, Charles W  
; REGISTRATION NUMBER: 28,946  
; REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; TELEX: 89-9456 LUKPAT  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-110-947-10  
Query Match 100.0%; Score 107; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 9.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 35 GETFOVEVPGSQHIDSOKKAI 55  
RESULT 3  
US-08-110-947A-26  
; Sequence 26. Application US/08110947A  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: COOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Matlare, Ltd.  
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson  
; STREET: Davis Highway  
; CITY: Arlington

STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,947A  
FILING DATE: 24-AUG-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLOW, Charles W  
REGISTRATION NUMBER: 28,946  
REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
TELEX: 89-9456 LUKPAT  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-110-947A-26  
Query Match 100.0%; Score 107; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 9.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 35 GETFOVEVPGSQHIDSOKKAI 55  
RESULT 4  
US-08-251-121-26  
; Sequence 26. Application US/08251121  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J  
; APPLICANT: STEIN, Penelope E  
; APPLICANT: COCKLE, Stephen A  
; APPLICANT: COOMEN, Raymond P  
; APPLICANT: KLEIN, Michel H  
; APPLICANT: ARMSTRONG, Glen D.  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/251,121  
; FILING DATE: 31-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-335

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-251-121-26

Query Match          100.0%; Score 107; DB 6; Length 93;
Best Local Similarity 100.0%; Pred. No. 9.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
Db 35 GETFOVEVPGSHIDSOKKAI 55

RESULT 5
US-08-150-305A-3
; Sequence 3, Application US/08150305A
; GENERAL INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; APPLICANT: Meinersmann, Richard J.
; TITLE OF INVENTION: CAMPYLOBACTER JEJUNI
; TITLE OF INVENTION: FLAGELLIN-ESCHERICHIA COLI ST-B FUSION PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 413, Building 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/150,305A
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0431.92
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-504-5676
; TELEFAX: 202-504-5060
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-150-305A-3

Query Match          100.0%; Score 107; DB 5; Length 371;
Best Local Similarity 100.0%; Pred. No. 5.5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
Db 49 GETFOVEVPGSHIDSOKKAI 69

RESULT 6
US-08-784-218-6
; Sequence 6, Application US/08784218
; GENERAL INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; APPLICANT: Meinersmann, Richard J.
; TITLE OF INVENTION: CAMPYLOBACTER JEJUNI
; TITLE OF INVENTION: FLAGELLIN-ESCHERICHIA COLI LT-B FUSION PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 411, Building 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,218
; FILING DATE: 16-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0043.97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-504-5676
; TELEFAX: 301-504-5060
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-784-218-6

Query Match          100.0%; Score 107; DB 11; Length 371;
Best Local Similarity 100.0%; Pred. No. 5.5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
Db 49 GETFOVEVPGSHIDSOKKAI 69

RESULT 7
US-08-829-026-5
; Sequence 5, Application US/08829026
; GENERAL INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; APPLICANT: Meinersmann, Richard J.
; TITLE OF INVENTION: CAMPYLOBACTER JEJUNI
; TITLE OF INVENTION: FLAGELLIN-ESCHERICHIA COLI LT-B FUSION PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janelle S. Graeter
; STREET: Room 411, Building 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,026
; FILING DATE: 18-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0106.97
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-504-5676  
TELEFAX: 301-504-5060  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 371 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-829-026-5

Query Match 100.0%; Score 107; DB 12; Length 371;  
Best Local Similarity 100.0%; Pred. No. 5.5e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHDSOKKAI 21  
|||||

Db 49 GETFOVEVPGSOHDSOKKAI 69

RESULT 8  
US-09-786-648-4  
Sequence 4, Application US/09786648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/GB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 4  
LENGTH: 21  
TYPE: PRT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 45...65  
OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivable f  
US-09-786-648-4

Query Match 94.4%; Score 101; DB 21; Length 21;  
Best Local Similarity 95.2%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHDSOKKAI 21  
|||||

Db 1 GATFOVEVPGSOHDSOKKAI 21

RESULT 9  
US-08-393-334-2  
Sequence 2, Application US/08393334  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M.  
APPLICANT: Yacoub, Reza K.  
APPLICANT: Zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression Of Gene Products From  
TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,334  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1135  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-393-334-2

Query Match 94.4%; Score 101; DB 7; Length 103;  
Best Local Similarity 95.2%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSOHDSOKKAI 21  
|||||

Db 45 GATFOVEVPGSOHDSOKKAI 65

RESULT 10  
US-08-782-832-15  
Sequence 15, Application US/08782832  
GENERAL INFORMATION:  
APPLICANT: Arntzen, Charles J.  
APPLICANT: Mason, Hugh S.  
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC  
TITLE OF INVENTION: BACTERIAL PROTEIN IN TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewilt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,832  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kettelberger, Denise M.  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 36170/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid



US-08-782-832-15

Matches 2

Db 45 GAT

US-08-817-906

APPLICANT

CORRESPONDENCE

STATE:

MEDIUM

COL 17  
CURRENT

PRIOR AP

NAME:

TELEPHONE  
TELETYPESEQUENCE  
LENGTH

US-08-817-906

**Matches** 2

Db 45 GAT

US-09-836-433

APPLICANT:

; CURRENT FI

; SEQ ID NO 14

US-09-836-433-

Matches	20
---------	----

Db 45 GATE

US-09-836-433-

APPLICANT:

1 CURRENT FILE

LENGTH: 11

CC-0 000 433

maladies 20

43 GATE  
DB

US-09-836-433 -  
: Sequence ??

APPLICANT:

; CURRENT FILE

; LENGTH: 11

; Sequence 55,



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:21:08 : Search time 47.5082 Seconds  
(without alignments)  
132.044 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEVPGSQHIDSOKKAI 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 1105779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	103	US-10-110-364-22	Sequence 22, Appl
2	107	100.0	123	US-10-110-364-20	Sequence 20, Appl
3	107	100.0	124	US-09-791-537-28360	Sequence 28360, A
4	107	100.0	124	US-09-791-537-29489	Sequence 29489, A
5	107	100.0	124	US-09-791-537-60743	Sequence 60743, A
6	107	100.0	124	US-10-110-364-16	Sequence 16, Appl
7	107	100.0	124	US-10-110-364-18	Sequence 18, Appl
8	107	100.0	124	US-10-110-364-23	Sequence 23, Appl
9	101	94.4	103	US-09-791-537-19387	Sequence 19387, A
10	101	94.4	103	US-09-791-537-38611	Sequence 38611, A
11	101	94.4	103	US-09-791-537-38639	Sequence 38639, A
12	101	94.4	103	US-09-791-537-74385	Sequence 74385, A
13	101	94.4	103	US-10-110-364-10	Sequence 10, Appl
14	101	94.4	103	US-10-110-364-13	Sequence 13, Appl
15	101	94.4	104	US-09-791-537-87980	Sequence 87980, A
16	101	94.4	104	US-09-791-537-99772	Sequence 99772, A
17	101	94.4	105	US-10-110-364-12	Sequence 12, Appl
18	101	94.4	113	US-09-791-537-73608	Sequence 73608, A
19	101	94.4	123	US-10-110-364-17	Sequence 17, Appl
20	101	94.4	124	US-09-791-537-40552	Sequence 40552, A
21	101	94.4	124	US-09-791-537-78640	Sequence 78640, A
22	101	94.4	124	US-09-791-537-92185	Sequence 92185, A
23	101	94.4	124	US-09-791-537-94644	Sequence 94644, A
24	101	94.4	124	US-09-791-537-103241	Sequence 103241, A
25	101	94.4	124	US-09-791-537-103660	Sequence 103660, A
26	101	94.4	124	US-10-110-364-2	Sequence 2, Appl

27	101	94.4	124	6	US-10-110-364-6	Sequence 6, Appl
28	101	94.4	124	6	US-10-110-364-11	Sequence 11, Appl
29	101	94.4	124	6	US-10-110-364-19	Sequence 19, Appl
30	101	94.4	124	6	US-10-110-364-21	Sequence 21, Appl
31	101	94.4	125	6	US-10-110-364-15	Sequence 15, Appl
32	101	94.4	129	5	US-09-791-537-131854	Sequence 131854, A
33	101	94.4	131	5	US-09-791-537-130348	Sequence 130348, A
34	101	94.4	138	6	US-10-141-6272-2	Sequence 2, Appl
35	95	88.8	103	5	US-09-791-537-68591	Sequence 68591, A
36	95	88.8	103	6	US-10-110-364-8	Sequence 8, Appl
37	95	88.8	124	5	US-09-791-537-33623	Sequence 33623, A
38	92	86.0	103	5	US-09-791-537-42610	Sequence 42610, A
39	92	86.0	103	6	US-10-110-364-7	Sequence 7, Appl
40	89	83.2	103	5	US-09-791-537-129309	Sequence 129309, A
41	89	83.2	103	6	US-10-110-364-5	Sequence 5, Appl
42	89	83.2	124	5	US-09-791-537-123948	Sequence 123948, A
43	89	83.2	124	6	US-10-110-364-4	Sequence 4, Appl
44	89	83.2	124	6	US-10-110-364-9	Sequence 9, Appl
45	89	83.2	382	1	PCT-US02-20978-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-10-110-364-22  
Sequence 22, Application US/10110364  
GENERAL INFORMATION:  
APPLICANT: Handley, Harold H.  
APPLICANT: Haaparanta, Tapio  
APPLICANT: Ewalt, Karla L.  
TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
FILE REFERENCE: ACTBIO. 004A  
CURRENT APPLICATION NUMBER: US/10/110,364  
CURRENT FILING DATE: 2002-04-05  
PRIOR APPLICATION NUMBER: PCT/US00/27607  
PRIOR FILING DATE: 2000-10-05  
PRIOR APPLICATION NUMBER: 60/158,561  
PRIOR FILING DATE: 1999-10-08  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 103  
TYPE: PRT  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)..(103)  
OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.  
US-10-110-364-22  
Query Match 100.0%; Score 107; DB 6; Length 103;  
Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GETFOVEVPGSQHIDSOKKAI 21  
|||||  
DB 45 GETFOVEVPGSQHIDSOKKAI 65  
RESULT 2  
US-10-110-364-20  
Sequence 20, Application US/10110364  
GENERAL INFORMATION:  
APPLICANT: Handley, Harold H.  
APPLICANT: Haaparanta, Tapio  
APPLICANT: Ewalt, Karla L.  
TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
FILE REFERENCE: ACTBIO. 004A  
CURRENT APPLICATION NUMBER: US/10/110,364  
CURRENT FILING DATE: 2002-04-05

```

; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 123
; TYPE: PRS
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(123)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.
US-10-110-364-20

Query Match
Best Local Similarity 100.0%; Score 107; DB 6; Length 123;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
Db 65 GETFOVEVPGSQHDSOKKAI 85

RESULT 3
US-09-791-537-28360
; Sequence 28360, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28360
; LENGTH: 124
; TYPE: PRS
; ORGANISM: Escherichia coli
US-09-791-537-28360

Query Match
Best Local Similarity 100.0%; Score 107; DB 5; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
Db 66 GETFOVEVPGSQHDSOKKAI 86

RESULT 4
US-09-791-29489
; Sequence 29489, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29489
; LENGTH: 124
; TYPE: PRS
; ORGANISM: Escherichia coli
US-09-791-537-29489
```

```

Query Match
Best Local Similarity 100.0%; Score 107; DB 5; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
Db 66 GETFOVEVPGSQHDSOKKAI 86

RESULT 5
US-09-791-537-60743
; Sequence 60743, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 60743
; LENGTH: 124
; TYPE: PRS
; ORGANISM: Escherichia coli
US-09-791-537-60743

Query Match
Best Local Similarity 100.0%; Score 107; DB 5; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
Db 66 GETFOVEVPGSQHDSOKKAI 86

RESULT 6
US-10-110-364-16
; Sequence 16, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 124
; TYPE: PRS
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1169505.
US-10-110-364-16

Query Match
Best Local Similarity 100.0%; Score 107; DB 6; Length 124;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GETFOVEVPGSQHDSOKKAI 21
Db 66 GETFOVEVPGSQHDSOKKAI 86
```

Db 66 GETFOVEVPGSQHIDSOKKAI 86

RESULT 7  
US-10-110-364-18

```
; Sequence 18, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 145833.
US-10-110-364-18
```

Query Match

100.0%; Score 107; DB 6; Length 124;  
Best Local Similarity 100.0%; Pred. No. 4.1e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 66 GETFOVEVPGSQHIDSOKKAI 86

RESULT 8  
US-10-110-364-23

```
; Sequence 23, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Escherichia coli
; NAME/KEY: VARIANT
; LOCATION: (1)...(124)
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 69630.
US-10-110-364-23
```

Query Match  
100.0%; Score 107; DB 6; Length 124;  
Best Local Similarity 100.0%; Pred. No. 4.1e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHIDSOKKAI 21

Db 66 GETFOVEVPGSQHIDSOKKAI 86

RESULT 9

```
US-09-791-537-19387
; Sequence 19387, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19387
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pbd 1CTID
US-09-791-537-19387
```

Query Match  
94.4%; Score 101; DB 5; Length 103;  
Best Local Similarity 95.2%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 10  
US-09-791-537-38611

```
; Sequence 38611, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38611
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pbd 1CHPD
US-09-791-537-38611
```

Query Match  
94.4%; Score 101; DB 5; Length 103;  
Best Local Similarity 95.2%; Pred. No. 3.4e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GETFOVEVPGSQHIDSOKKAI 21  
Db 45 GATFOVEVPGSQHIDSOKKAI 65

RESULT 11

```
US-09-791-537-38639
; Sequence 38639, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
```



Fri Oct 25 17:51:35 2002

us-09-786-648-5.rapn

Page 5

```
Qy 1 GETFOVEVPGSQHIDSQKAI 21
    | | | | | | | | | |
Db 46 GATFOVEVPGSQHIDSQKAI 66
```

Search completed: October 24, 2002, 15:33:20  
Job time : 48.5082 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:17:58 : Search time 16.1803 Seconds

(without alignments)  
124.712 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEVPGSQHIDSOKKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	124	1 QLECB	heat-labile entero
2	95	88.8	124	1 XVVCB	cholera enterotoxi
3	49	45.8	255	2 A86457	probable peptide c
4	47	43.9	392	2 T04150	RAD3 protein homo
5	46	43.0	91	2 C96580	hypothetical prote
6	45	43.0	574	2 C86400	hypothetical prote
7	45	42.1	439	2 T49189	kinesin heavy chai
8	45	42.1	2733	2 S15760	genome polyprotein
9	44	41.1	263	2 S78364	conserved hypotet
10	44	41.1	374	2 T19866	hypothetical prote
11	44	41.1	864	2 T49574	probable carnitine
12	44	41.1	2731	1 VETJHJ	genome polyprotein
13	43.5	40.7	641	2 C84726	probable receptor-
14	43	40.2	128	1 UDDOR	ubiquitin / riboso
15	43	40.2	154	1 UDDOR7	ubiquitin / riboso
16	43	40.2	159	2 A97466	hypothetical prote
17	43	40.2	159	2 AC2684	hypothetical prote
18	43	40.2	228	2 D34080	ubiquitin 18 - S11
19	43	40.2	229	2 B27806	ubiquitin (clone 1
20	43	40.2	368	2 T04861	hypothetical prote
21	43	40.2	380	2 B34080	polyubiquitin 5 (c
22	43	40.2	380	2 B34080	polyubiquitin 5 (c
23	43	40.2	381	2 A27806	polyubiquitin 5 (c
24	43	40.2	532	2 A34080	polyubiquitin 7 (c
25	43	40.2	1810	1 A32230	tenascin precursor
26	42	39.3	304	2 E82963	probable peptide c
27	42	39.3	330	2 C89848	peptide chain rele
28	42	39.3	427	2 JC5694	stress-activated p
29	42	39.3	672	2 A87441	penicillin-binding

30	42	39.3	1742	2 S76110	hypothetical prote
31	42	39.3	1959	2 AG1085	hypothetical prote
32	41.5	38.8	376	1 S17246	chondriate synthas
33	41.5	38.8	500	2 JC4022	4-aminobutyrate tr
34	41	38.3	166	2 G90661	probable peptide c
35	41	38.3	166	2 E64748	translation releas
36	41	38.3	166	2 E85512	probable peptide c
37	41	38.3	204	2 A10542	probable peptide c
38	41	38.3	260	2 C96827	protein F20B17.2 f
39	41	38.3	269	1 A25973	pertussis toxin ch
40	41	38.3	313	2 E88216	protein B0495.8 (1
41	41	38.3	333	2 D87512	alcohol dehydrogen
42	41	38.3	386	1 S66056	yaan protein - Bac
43	41	38.3	399	2 T46898	queuing tRNA-ribos
44	41	38.3	550	2 T01770	hypothetical prote
45	41	38.3	683	2 S01433	repressor protein

## ALIGNMENTS

RESULT 1  
QLECB  
heat-labile enterotoxin chain B precursor - Escherichia coli  
C:Species: Escherichia coli  
C>Date: 29-Jun-1981 #sequence-revision 29-Jun-1981 #text-change 18-Jun-1999  
C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
R:Dallas, W.S.; Falkow, S.  
Nature 288, 499-501, 1980  
A>Title: Amino acid sequence homology between cholera toxin and Escherichia coli heat  
A:Reference number: A01820; MUID:81074965  
A:Accession: A01820  
A:Molecule type: mRNA  
A:Residues: 1-124 <DAL>  
R:Yamamoto, T.; Gojibori, T.; Yokota, T.  
J. Bacteriol. 169, 1352-1357, 1987  
A>Title: Evolutionary origin of pathogenic determinants in enterotoxigenic Escherich  
A:Reference number: A26946; MUID:87137303  
A:Accession: B26946  
A:Molecule type: DNA  
A:Residues: 1-27, 'E', 29-63, 'K', 65-124 <YAM>  
A:Cross-references: EMBL:M15363; NID:g148335; PIDN:AAA24792.1; PID:g148336  
R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
Infect. Immun. 48, 73-77, 1985  
A>Title: Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons  
A:Reference number: I41194; MUID:85156481  
A:Accession: I41194  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5, 'F', 7-17, 'C', 19-24, 'S', 26-27, 'E', 29-33, 'H', 35-63, 'K', 65-66, 'A', 68-122  
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376  
A:Experimental source: plasmid ENT-R PCG86  
R:Ibrahim, I.; Gentz, R.  
J. Biol. Chem. 262, 10189-10194, 1987  
A>Title: A functional interaction between the signal peptide and the translation appa  
titulum.  
A:Reference number: I41287; MUID:87280041  
A:Accession: I41287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-22 <RE2>  
A:Cross-references: GB:M17101; NID:g146375; PIDN:AAA23973.1; PID:g146376  
FEMS Microbiol. Lett. 108, 157-161, 1993  
A>Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
A:Reference number: I53542; MUID:93252225  
A:Accession: I67644  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-17, 'C', 19, 'Y', 21-24, 'S', 26-27, 'E', 29-63, 'K', 65-66, 'A', 68-122, 'E', 124 <R  
A:Cross-references: GB:S60731; NID:g1408994; PIDN:AA650441.1; PID:g1408996  
R:Tsuji, T.; Lida, T.; Honda, T.; Miwatani, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
Microb. Pathog. 2, 381-390, 1987







Rilee, H.J.; Shieh, C.K.; Gorbatenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagd submitted to GenBank, February 1991  
A:Description: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding  
A:Reference number: A36815  
A:Accession: B36815  
A:Molecule type: genomic RNA  
A:Residues: 1-2731 <LEP>  
A:Cross-references: GB:M55148; NID:9331851; PIDN:AAA6458.1; PID:9331853  
Rilee, H.J.; Shieh, C.K.; Gorbatenya, A.E.; Koonin, E.V.; La Monica, N.; Tuler, J.; Bagd  
Virology 180, 567-582, 1991  
A:Title: The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the  
A:Reference number: A38547; MUID:91111976  
A:Contents: annotation  
A:Note: neither nucleotide nor complete amino acid sequence is given  
C:Comment: This protein may be translated as a 1a-1b polypeptide by a ribosomal frameshift  
C:Genetics:  
A:Gene: 1b  
C:Superfamily: Infectious bronchitis virus RNA-directed RNA polymerase  
C:Keywords: glycoprotein; nucleotidyltransferase; RNA biosynthesis  
F:269,304,785,1184,1287,1524,1842,2196,2575,2630,2645,2665/Binding site: carbohydrate (R  
Query Match 41.1%; Score 44; DB 1; Length 2731;  
Best Local Similarity 69.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 EFMFVEVPGSOH 14  
DB 1198 EFMFNNVPMYOH 1210  
RESULT 13  
C84726  
probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84726  
Rilin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umeyam, L.; Tallon, L.;  
euss, D.; Niernan, M.C.; White, O.; Eisen, J.A.; Salberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: C84726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-641 <STO>  
A:Cross-References: GB:AE002093; NID:g4887748; PIDN:AAD32284.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31880  
A:Map position: 2  
Query Match 40.7%; Score 43.5; DB 2; Length 641;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
QY 1 GEFVEVPGSQ-HIDSOKKAI 21  
DB 359 GEVFKALPGSNGKITAVKKVI 380  
RESULT 14  
UDOR  
ubiquitin / ribosomal protein CEP52 - slime mold (Dictyostelium discoideum)  
N:Alternate names: ubiquitin fusion protein  
C:Species: Dictyostelium discoideum  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
C:Accession: S00357; A25863  
R.Mueller-Taubenberger, A.; Westphal, M.; Jaeger, E.; Noegel, A.; Gerisch, G.  
FEBS Lett. 229, 273-278, 1988  
A:Title: Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-terminal tail  
A:Reference number: S00357; MUID:88152253  
A:Accession: S00357  
A:Molecule type: mRNA

A:Residues: 1-128 <MUE>  
A:Cross-References: EMBL:X07210; NID:g7381; PIDN:CAA30183.1; PID:g7382  
A:Experimental source: strain AX2-214  
C:Genetics:  
A:Gene: DUB1  
C:Superfamily: ubiquitin/ribosomal protein CEP52; ribosomal protein CEP52 homology; u  
C:Keywords: DNA binding; protein biosynthesis; protein degradation; ribosome; zinc fi  
F:1-76/Product: ubiquitin #status predicted <DB1>  
F:1-76/Domain: ubiquitin homology <UBH>  
F:77-128/Product: ribosomal protein CEP52 #status predicted <RIB>  
F:77-128/Domain: ribosomal protein CEP52 homology <CPH>  
F:95-114/Region: zinc finger CCCC motif  
F:121-128/Region: nuclear location signal  
Query Match 40.2%; Score 43; DB 1; Length 128;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 1 GEFVEVPGSOHIDSOKKAI 21  
DB 10 GKTITLEVEGSDNIENYKAKI 30  
RESULT 15  
UDOR7  
ubiquitin / ribosomal protein S27a - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 21-Jul-2000  
C:Accession: E34080  
R.Ohmachi, T.; Giorda, R.; Shaw, D.R.; Ennis, H.L.  
Biochemistry 28, 5226-5231, 1989  
A:Title: Molecular organization of developmentally regulated Dictyostelium discoideum  
A:Reference number: A34080; MUID:89352609  
A:Accession: E34080  
A:Molecule type: mRNA  
A:Residues: 1-154 <OHM>  
A:Cross-References: GB:M23750; GB:J02858; NID:g167940; PIDN:AAA33264.1; PID:g167941  
C:Superfamily: ubiquitin / rat ribosomal protein S27a; ribosomal protein S27a homology  
C:Keywords: protein biosynthesis; protein degradation; ribosome  
F:1-76/Product: ubiquitin #status predicted <MAT1>  
F:1-76/Domain: ubiquitin homology <UBH>  
F:77-154/Product: ribosomal protein S27a #status predicted <MAT2>  
F:102-152/Domain: ribosomal protein S27a homology <RIB>  
Query Match 40.2%; Score 43; DB 1; Length 154;  
Best Local Similarity 42.9%; Pred. No. 17;  
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 1 GEFVEVPGSOHIDSOKKAI 21  
DB 10 GKTITLEVEGSDNIENYKAKI 30  
Search completed: October 24, 2002, 15:23:24  
Job time : 18.1803 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:42 ; Search time 8.26229 Seconds  
(without alignments)  
98.412 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFQVEVPGSHIDSQKKAI 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	124	1 ELBP_ECOLI	P32890 escherichia
2	101	94.4	124	2 ELBH_ECOLI	P13811 escherichia
3	95	88.8	124	1 CHTB_VIBCH	P01556 vibrio chol
4	45	42.1	2733	1 RRPB_CYMA5	P16342 murine coro
5	44	41.1	263	1 YC43_ODOST	P49538 odonella s
6	44	41.1	467	1 IRP6_HUMAN	O14896 homo sapien
7	44	41.1	467	1 IRP6_MOUSE	P97431 mus musculi
8	44	41.1	2731	1 RRPB_CYMAJ	P29882 murine coro
9	43	40.2	76	1 UBIO_DICDI	P08618 dictyostell
10	43	40.2	1808	1 TENA_CHICK	P10039 gallus gall
11	42	39.3	855	1 ST14_HUMAN	O99376 homo sapien
12	41.5	38.8	376	1 AROC_YEAST	P28777 saccharomyc
13	41.5	38.8	500	1 GABT_HUMAN	P80404 homo sapien
14	41	38.3	141	1 RPH_ECOLI	P28369 escherichia
15	41	38.3	313	1 YP68_CAEEL	O09217 caenorhabdi
16	41	38.3	385	1 TGT_ZYMO	P28720 zymomonas m
17	41	38.3	386	1 YTAAN_BACSU	P37535 bacillus su
18	41	38.3	567	1 STP_STRCO	O94796 streptomyce
19	41	38.3	683	1 RPC_BPPHC	P08799 bacterioph
20	41	38.3	1702	1 DPOL_THELI	P30317 thermococcu
21	41	38.3	1829	1 DPOL_THEST	O33845 thermococcu
22	40.5	37.9	862	1 PMS2_HUMAN	P54278 homo sapien
23	40	37.4	76	1 UBIO_NEUCR	P13117 neuropeptid
24	40	37.4	148	1 IF5A_AERPE	O99453 aeropyrum p
25	40	37.4	288	1 HS74_CAEEL	P20163 caenorhabdi
26	40	37.4	309	1 CTR2_RAT	O92419 rattus norv
27	40	37.4	355	1 YRY1_CAEEL	O10005 caenorhabdi
28	40	37.4	395	1 ADH3_CENTH	O24577 entomobda h
29	40	37.4	523	1 DBP3_YEAST	P20447 saccharomyc
30	40	37.4	773	1 DPOL_THES9	P56689 thermococcu
31	40	37.4	775	1 DPOL_THES9	O56366 thermococcu
32	40	37.4	1048	1 CPXB_BACME	P14779 bacillus me
33	40	37.4	1523	1 DPOL_THEFM	P74918 thermococcu

34	40	37.4	1668	1 DPOL_THEHY	O9h05 thermococcu
35	40	37.4	1671	1 DPOL_PYRKO	P77933 pyrococcus
36	40	37.4	1693	1 DPOL_THES8	O9h84 thermococcu
37	39.5	36.9	1090	1 PULA_KLEPN	P07206 klebsiella
38	39	36.4	76	1 UBIO_YEAST	P04838 saccharomyc
39	39	36.4	271	1 CEAM_ECOLI	P05820 escherichia
40	39	36.4	320	1 Y054_MYCPN	P75049 mycoplasma
41	39	36.4	445	1 RF1M_HUMAN	O75570 homo sapien
42	39	36.4	456	1 SR54_THESAC	O9hK0 thermoplasm
43	39	36.4	459	1 IL7R_MOUSE	P16872 mus musculi
44	39	36.4	491	1 CD5_RAT	P51882 rattus norv
45	39	36.4	560	1 INRI_SHEEP	O28589 ovis aries

## ALIGNMENTS

RESULT 1	ID	ELBP_ECOLI	STANDARD:	PRT:	124 AA.
AC	P32850;	P13768;	P01557;		
DT	21-JUL-1986	(Rel. 01, Created)			
DT	01-NOV-1995	(Rel. 32, last sequence update)			
DT	15-JUL-1998	(Rel. 36, last annotation update)			
DE	Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTP-B).				
GN	ELTB OR LTPB.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OX	NCBI_TaxId=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ISOLATE P307;				
RX	MEDLINE=81074965; PubMed=7003397;				
RA	Dallas W.S., Falkow S.;				
RT	"Amino acid sequence homology between cholera toxin and Escherichia				
RT	coli heat-labile toxin.";				
RL	Nature 288:499-501(1980).				
RN	[2]				
RP	REVISIONS TO 28 AND 64.				
RC	STRAIN-ISOLATE P307;				
RX	MEDLINE=85156481; PubMed=3884513;				
RA	Leong J., Vinal A.C., Dallas W.S.;				
RT	"Nucleotide sequence comparison between heat-labile toxin B-subunit				
RT	cistrons from Escherichia coli of human and porcine origin.";				
RL	Infect. Immun. 48:73-77(1985).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ISOLATE PCG86;				
RX	MEDLINE=87137303; PubMed=3546273;				
RA	Yamamoto T., Gojodori T., Yokota T.;				
RT	"Evolutionary origin of pathogenic determinants in enterotoxigenic				
RT	Escherichia coli and Vibrio cholerae O1.";				
RL	J. Bacteriol. 169:1352-1357(1987).				
RN	[4]				
RP	SEQUENCE OF 1-22 FROM N.A.				
RX	MEDLINE=87280041; PubMed=3301830;				
RA	Ibrahim I., Gentz R.;				
RT	"A functional interaction between the signal peptide and the				
RT	translation apparatus is detected by the use of a single point				
RT	mutation which blocks translocation across mammalian endoplasmic				
RL	reticulum.";				
RN	J. Biol. Chem. 262:10189-10194(1987).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).				
RX	MEDLINE=93240541; PubMed=8478941;				
RA	Sigma T. K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;				
RT	"Refined structure of Escherichia coli heat-labile enterotoxin, a				
RT	close relative of cholera toxin.";				
RL	J. Mol. Biol. 230:890-918(1993).				
RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).				
RX	MEDLINE=91238966; PubMed=2034287;				

RA Sijma T.K., Pronk S.E., Kalk K.H., Martina E.S., van Zanten B.A.M.,  
 RA Witholt B., Hol W.G.J.:  
 RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin  
 RT from *E. coli*.";  
 RL Nature 351:371-377(1991).  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE-95349400; PubMed-7623669;  
 RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappuoli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of *Escherichia coli* and *Vibrio cholerae*.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M17873; AAA98065.1; -;  
 DR EMBL; M15363; AAA24792.1; -;  
 DR EMBL; M17101; AAA23973.1; -;  
 DR PIR; A01820; OLECB.  
 DR PIR; B26946; OLECEB.  
 DR PDB; 1LTA; 31-JAN-94.  
 DR PDB; 1LTB; 31-JAN-94.  
 DR PDB; 1LTG; 15-SEP-95.  
 DR PDB; 1LTI; 17-AUG-96.  
 DR PDB; 1LTS; 31-JAN-94.  
 DR PDB; 1LTT; 31-JAN-94.  
 DR PDB; 1LTJ; 07-JUL-97.  
 DR PDB; 1LTK; 16-JUN-97.  
 DR PDB; 1LTI; 03-DEC-97.  
 DR PDB; 1LTG; 03-DEC-97.  
 DR PDB; 1LTI; 20-APR-95.  
 DR InterPro: IPR001835; Enterotoxin\_B.  
 DR Pfam: PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR ProDom; PD012805; Enterotoxin\_B; 1.  
 KW Enterotoxin; Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT STRAND 36 43  
 FT STRAND 47 51  
 FT TURN 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT HELIX 80 98  
 FT TURN 99 100  
 FT STRAND 103 109  
 FT STRAND 115 123  
 SQ SEQUENCE 124 AA; 14133 MW; 6DB/DE58395EA70D CRC64;  
 Query Match 100.0%; Score 107; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 3e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
 ID EMBL\_ECOLI STANDARD; PRT; 124 AA.  
 AC P13811;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).  
 GN ELTB OR LTBP.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H74-114;  
 RX MEDLINE-85156481; PubMed-3884513;  
 RA Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT cistrons from *Escherichia coli* of human and porcine origin.";  
 RL Infect. Immun. 48:73-77(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H10407;  
 RX MEDLINE-83114628; PubMed-6759877;  
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
 RT "Overlapping genes in the heat-labile enterotoxin operon originating  
 RT from *Escherichia coli* human strain.";  
 RL Mol. Gen. Genet. 188:356-359(1982).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H10407;  
 RX MEDLINE-93252225; PubMed-8486242;  
 RA Inoue T., Tsuji T., Koto M., Imanura S., Miyama A.;  
 RT "Amino acid sequence of heat-labile enterotoxin from chicken  
 RT enterotoxigenic *Escherichia coli* is identical to that of human strain  
 RT H 10407.";  
 RL FEBS Microbiol. Lett. 108:157-161(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ETEC LT 87;  
 RA Germain Y., Desperrier J.M.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE-95349400; PubMed-7623669;  
 RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappuoli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of *Escherichia coli* and *Vibrio cholerae*.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
 RX MEDLINE-99185101; PubMed-10085117;  
 RA Matkovic-Calogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,  
 RA Tossi A., Palu G., Zanotti G.;  
 RT "Crystal structure of the B subunit of *Escherichia coli* heat-labile  
 RT enterotoxin carrying peptides with anti-herpes simplex virus type 1  
 RT activity.";  
 RL J. Biol. Chem. 274:8764-8769(1999).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M17874; AAA98064.1; -;



DR EMBL: J01646; AAB02982.1; -  
 DR EMBL: S60731; AAC60441.1; -  
 DR EMBL: X83966; CAA58800.1; -  
 DR PDB: 1LTR; 23-MAR-99.  
 DR InterPro: IPR001835; Enterotoxin\_B.  
 DR Pfam: PF01376; Enterotoxin\_B; 1.  
 DR PRINTS: PR00772; ENTEROTOXINB.  
 DR ProDom: PD012805; Enterotoxin\_B; 1.  
 KW Enterotoxin; Signal: 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT VARIANT 6 6 F -> C (IN ISOLATE H10407).  
 FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).  
 FT VARIANT 34 34 H -> R (IN ISOLATE H10407).  
 SQ SEQUENCE 124 AA; 14027 MW; E9F7F7C7B9D3BC47 CRC64;  
 Query Match 94.4%; Score 101; DB 1; Length 124;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 GETFQVEVPGSHIDSQKAI 21  
 Db 66 GATFQVEVPGSHIDSQKAI 86  
 RESULT 3  
 CHTB\_VIBCH STANDARD: PRT; 124 AA.  
 AC P01556; Q9J002;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cholera enterotoxin, beta chain precursor.  
 GN CTXB OR TOXB OR VC1456.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84061784; PubMed=6315707;  
 RA Lockman H., Kaper J.B.;  
 RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio  
 cholerae enterotoxin."  
 RL J. Biol. Chem. 258:13722-13726(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=EL TOR 2125;  
 RC MEDLINE=84068199; PubMed=6646234;  
 RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,  
 de Wilde M.;  
 RT "Cholera toxin genes: nucleotide sequence, deletion analysis and  
 vaccine development."  
 RL Nature 306:551-557(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR 2125;  
 RA Dams E., de Wolf M., Dierick W.;  
 RT Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=4260B / SEROTYPE O139;  
 RC MEDLINE=94237453; PubMed=8181723;  
 RA Lebens M., Holmgren J.;  
 RT "Structure and arrangement of the cholera toxin genes in Vibrio  
 cholerae O139."  
 RL FEMS Microbiol. Lett. 117:197-202(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1854 / O139-BENGAL;  
 RA Yamamoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,  
 Honda T.;  
 RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.

RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae."  
 RL Nature 406:477-483(2000).  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005357; PubMed=903363;  
 RA Kurosky A., Markel D.E., Peterson J.W.;  
 RT "Covalent structure of the beta chain of cholera enterotoxin."  
 RL J. Biol. Chem. 252:7257-7264(1977).  
 RN [8]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005356; PubMed=903362;  
 RA Lai C.-Y.;  
 RT "Determination of the primary structure of cholera toxin B subunit."  
 RL J. Biol. Chem. 252:7249-7256(1977).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=94272319; PubMed=8003954;  
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martial J.A.,  
 Hol W.G.J.;  
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
 pentasaccharide."  
 RL Protein Sci. 3:166-175(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=95387394; PubMed=7658472;  
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,  
 Maulik P.R., Reed R.A., Shipley G.G.;  
 RT "The 2.4 Å crystal structure of cholera toxin B subunit pentamer:  
 choleraenoid."  
 RL J. Mol. Biol. 251:550-562(1995).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RC STRAIN=OGAWA 41 / CLASSICAL BIOTYPE;  
 RX MEDLINE=97376625; PubMed=9232653;  
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,  
 Hirst T.R., Hol W.G.J.;  
 RT "Structural studies of receptor binding by cholera toxin mutants."  
 RL Protein Sci. 6:1516-1528(1997)  
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
 BINDING TO CELL MEMBRANES.  
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN  
 (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
 DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
 6 BETA CHAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X00171; CAA24996.1; -  
 DR EMBL: K01170; AAA27573.1; -  
 DR EMBL: D30053; BAA06291.1; -  
 DR EMBL: X58786; CAA1593.1; -  
 DR EMBL: X76390; CAA53973.1; -  
 DR EMBL: X76391; CAA53976.1; -  
 DR EMBL: AE004224; AAF94613.1; -  
 DR PIR: A01819; XVVCB.



DR PROSITE; PS01218; TATC; 1.  
 KW Chloroplast; Hypothetical protein; Transmembrane.  
 FT TRANSMEM 53 POTENTIAL.  
 FT TRANSMEM 103 123 POTENTIAL.  
 FT TRANSMEM 130 150 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 213 233 POTENTIAL.  
 FT TRANSMEM 241 261 POTENTIAL.  
 SQ SEQUENCE 263 AA; 30080 MW; 25083364F2A89E88 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 263;  
 Best Local Similarity 40.0%; Pred. No. 8.3;  
 Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

OY 2 ETVFVPSGSHIDSOKKAI 21  
 DB 30 ETVVTELPFSEHTEELKORL 49

## RESULT 6

IRF6\_HUMAN  
 ID IRF6\_HUMAN STANDARD; PRT; 467 AA.  
 AC O14896;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Interferon regulatory factor 6 (IRF-6).  
 GN IRF6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grossman A., Mitrucker H.W., Antonio L., Oato K., Mak T.W.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RM [2]  
 RP SEQUENCE FROM N.A.  
 RA Graham D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RM [2]  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; AF021292; AAB84111.1; -;  
 DR EMBL; AL022398; CA18545.1; -;  
 DR HSSP; P23906; 2IRF.  
 DR InterPro: IPR001346; IRF.  
 DR Pfam: PF00605; IRF; 1.  
 DR PRINTS; PR00267; INTERPREGCT.  
 DR ProDom; PD002355; IRF; 1.  
 DR SMART; SM00348; IRF; 1.  
 DR PROSITE; PS00601; IRF; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.  
 SQ SEQUENCE 467 AA; 53129 MW; 7E28F5E0F5BA4053 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 467;  
 Best Local Similarity 41.2%; Pred. No. 15;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 5 QVEVPSGSHIDSOKKAI 21  
 DB 273 QVKEPGEHTEHTEKORL 289

RESULT 7  
 IRF6\_MOUSE  
 ID IRF6\_MOUSE STANDARD; PRT; 467 AA.  
 AC P97431;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Interferon regulatory factor 6 (IRF-6).  
 GN IRF6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BALB/C; TISSUE-Colon;  
 RA Grossman A., Mitrucker H.W., Antonio L., Mak T.W.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RM [2]  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

CC EMBL; U73029; AAB36714.1; -;  
 DR HSSP; P23906; 2IRF.  
 DR MGD; MGI:1859211; Irf6.  
 DR InterPro: IPR001346; IRF.  
 DR Pfam; PF00605; IRF; 1.  
 DR PRINTS; PR00267; INTERPREGCT.  
 DR ProDom; PD002355; IRF; 1.  
 DR SMART; SM00348; IRF; 1.  
 DR PROSITE; PS00601; IRF; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.  
 SQ SEQUENCE 467 AA; 53106 MW; 68CCAA90680FEDC8 CRC64;

Query Match 41.1%; Score 44; DB 1; Length 467;  
 Best Local Similarity 41.2%; Pred. No. 15;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 5 QVEVPSGSHIDSOKKAI 21  
 DB 273 QVKEPGEHTEHTEKORL 289

## RESULT 8

RRPB\_CVMJH  
 ID RRPB\_CVMJH STANDARD; PRT; 2731 AA.  
 AC P29982;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (ORF1b).  
 DE Murine coronavirus MHV (strain JHM).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxId=11144;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-91111976; PubMed-1846489;  
 RA Lee H.-Y., Shieh C.-K., Gorbaleya A.E., Koonin E.V., la Monica N.,  
 RA Tulen J., Bagdazhardyan A., Lai M.W.C.;  
 RT "The complete sequence (22 kilobases) of murine coronavirus gene 1  
 RT encoding the putative proteases and RNA polymerase.";  
 RL Virology 180:567-582(1991).  
 CC -1- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS

CC A MULTIFUNCTIONAL PROTEIN. IT CONTAINS THE ACTIVITIES NECESSARY  
 CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,  
 CC SUBGENOMIC MRNAS AND PROGENY VIRION RNA.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNAI)(N).  
 CC -1- MISCELLANEOUS: THIS PROTEIN MAY BE TRANSLATED AS A 1A-1B  
 CC POLYPROTEIN BY A RIBOSOMAL FRAMESHIFTING MECHANISM.  
 CC -1- SIMILARITY: TO THE AVIAN CORONAVIRUS IBV RNA POLYMERASE.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M55148; AAA46458.2; -  
 DR PIR: B36815; VFIHJH.  
 KM Transferase: RNA-directed RNA polymerase; Helicase; ATP-binding.  
 FT DOMAIN 622 824 POLYMERASE.  
 FT DOMAIN 944 1014 CYS/HIS-RICH.  
 FT DOMAIN 1218 1506 HELICASE.  
 FT NP\_BIND 1220 1227 ATP (BY SIMILARITY).  
 SQ SEQUENCE 2731 AA; 308852 MW; 99463066B176970 CRC64;  
 Query Match 41.18; Score 44; DB 1; Length 2731;  
 Best Local Similarity 69.28; Pred. No. 1e+02;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Oy 2 ETPOVEVPGSOHI 14  
 Db 1198 ETPONVNPYQHI 1210

RESULT 9  
 UBIQ\_DICDI STANDARD; PRT; 76 AA.  
 ID UBIQ\_DICDI STANDARD; PRT; 76 AA.  
 AC P08618;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-NOV-1991 (Rel. 20, Last annotation update)  
 DE Ubiquitin.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8152253; PubMed=2831095;  
 RA Mueller-Taubenberger A., Westphal M., Jaeger E., Noegel A.,  
 RA Gerisch G.;  
 RT "Complete cDNA sequence of a Dictyostelium ubiquitin with a carboxy-  
 RT terminal tail and identification of the protein using an anti-peptide  
 RT antibody";  
 RL FEBS Lett. 229:273-278(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89352609; PubMed=2548604;  
 RA Omachail T., Giorada R., Shaw D.R., Ennis H.L.;  
 RA "Molecular organization of developmentally regulated Dictyostelium  
 RT discoideum ubiquitin cDNAs";  
 RL Biochemistry 28:5226-5231(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87257921; PubMed=3037345;  
 RA Giorada R., Ennis H.L.;  
 RT "Structure of two developmentally regulated Dictyostelium discoideum  
 RT ubiquitin genes";  
 RL Mol. Cell. Biol. 7:2097-2103(1987).  
 RN [4]  
 RP SEQUENCE OF 13-76 FROM N.A.  
 RC STRAIN=AX2;  
 RA Westphal M., Mueller-Taubenberger A., Noegel A., Gerisch G.;

RT "Transcript regulation and carboxyterminal extension of ubiquitin in  
 RT Dictyostelium discoideum";  
 RL FEBS Lett. 209:92-96(1986).  
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SELECTIVE DEGRADATION OF  
 CC CELLULAR PROTEINS, THE MAINTENANCE OF CHROMATIN STRUCTURE, THE  
 CC REGULATION OF GENE EXPRESSION, THE STRESS RESPONSE, AND RIBOSOME  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
 CC -1- MISCELLANEOUS: UBIQUITIN IS SYNTHESIZED AS A POLYUBIQUITIN  
 CC PRECURSOR WITH EXACT HEAD TO TAIL REPEATS. SOME UBIQUITIN GENES  
 CC CONTAIN A SINGLE COPY OF UBIQUITIN FUSED TO A RIBOSOMAL PROTEIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07210; CAA30183.1; ALT\_TERM.  
 DR EMBL: M19666; AAA33261.1; ALT\_TERM.  
 DR EMBL: M19491; AAA33269.1; ALT\_TERM.  
 DR EMBL: M19492; AAA33270.1; ALT\_TERM.  
 DR EMBL: M23748; AAA33262.1; ALT\_TERM.  
 DR EMBL: M23749; AAA33263.1; ALT\_TERM.  
 DR EMBL: M23750; AAA33264.1; ALT\_TERM.  
 DR EMBL: M23751; AAA33265.1; ALT\_TERM.  
 DR EMBL: M23752; AAA33266.1; ALT\_TERM.  
 DR EMBL: M23753; AAA33267.1; ALT\_TERM.  
 DR EMBL: M23754; AAA33268.1; ALT\_TERM.  
 DR EMBL: X04702; CAA28408.1; ALT\_TERM.  
 DR PIR: S00357; UQDOR.  
 DR PIR: E34080; UQDOR7.  
 DR PIR: A27806; A27806.  
 DR PIR: B27806; B27806.  
 DR PIR: A34080; A34080.  
 DR PIR: B34080; B34080.  
 DR PIR: C34080; C34080.  
 DR PIR: D34080; D34080.  
 DR HSSP: P02248; UBI.  
 DR Dictydb: DP05001; ubqA.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR PRINTS: PR00348; UBIQUITIN.  
 DR SMART: SM00213; UBO; 1.  
 DR PROSITE: PS00299; UBIQUITIN\_1; 1.  
 DR PROSITE: PS50053; UBIQUITIN\_2; 1.  
 DR Nuclear protein; Polyprotein.  
 FT SITE 48  
 FT NUCLEAR PROTEIN; POLYPROTEIN.  
 FT BINDING 76  
 FT VARIANT 11 76  
 SQ SEQUENCE 76 AA; 8538 MW; 642738396BEA84 CRC64;  
 Query Match 40.28; Score 43; DB 1; Length 76;  
 Best Local Similarity 42.98; Pred. No. 3.2;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
 Oy 1 GETFOVEVPGSOHIDSOAKI 21  
 Db 10 KRTTLEVEGSDNIENYAKKI 30

RESULT 10  
 TENA\_CHICK STANDARD; PRT; 1808 AA.  
 ID TENA\_CHICK STANDARD; PRT; 1808 AA.  
 AC P10039; P13132; O73584; O73585;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tenascin precursor (TN) (Hexabrachion) (Cytotactin) (Neurexectin)  
 DE (GHEW) (JT) (Motendinous antigen) (Glioma-associated-extracellular



```

FT DISULFID 506 517 BY SIMILARITY.
FT DISULFID 519 528 BY SIMILARITY.
FT DISULFID 533 543 BY SIMILARITY.
FT DISULFID 537 548 BY SIMILARITY.
FT DISULFID 550 559 BY SIMILARITY.
FT DISULFID 564 574 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 581 590 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1101 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1112 1112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1153 1153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1736 1736 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 1043 1224 MISSING (IN 200 KDA ISOFORM).
FT VARSPIC 1043 1315 MISSING (IN 190 KDA ISOFORM).
FT CONFLICT 182 182 W -> R (IN REF. 2).
FT CONFLICT 563 571 SCPDNNV -> PAPMTATW (IN REF. 3).
FT CONFLICT 598 598 E -> G (IN REF. 3).
FT CONFLICT 838 838 T -> TXY (IN REF. 3).
FT CONFLICT 886 886 N -> F (IN REF. 3).
SQ SEQUENCE 1808 AA: 19858 MW: B924AD0CF9EFD6DE CRC64:

Query Match 40.2%; Score 43; DB 1; Length 1808;
Best Local Similarity 58.3%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ETOVEYVPGSOH 13
Db 1087 ETWNTVPGSOH 1098
11: : 111 11

```

```

RN [3]
RP SEQUENCE FROM N.A.
RP TISSUE=Prostate;
RA Yamaguchi N., Mitsui S.;
RT "Molecular cloning of a novel transmembrane serine protease expressed
in human prostate.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
O'Brien T.J.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 327-855 FROM N.A.
RP TISSUE=Muscle;
RA Strausberg R.;
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SMC19.";
RN [7]
RP submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilysin
and a Kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -I- FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
OR LYS AS THE P1 SITE.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -I- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: AF118224; AAD42765.2; -
CC EMBL: AF133086; AAF00109.1; -
CC EMBL: AB030036; BAB20376.1; -
CC EMBL: AF057145; AAG15395.1; -
CC EMBL: BC005826; AAH05826.1; -
CC EMBL: AF283256; AAG13949.1; -
CC HSP: P00763; IDPO.
CC MEROPS: S01.302; -.
CC InterPro: IPR000859; CUB.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR002172; LDL_recept_A.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF00057; Idl_recept_a; 4.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PRO0722; CHYMOTRYPSIN.
CC PRINTS: PRO0261; LDLRECEPTOR.
CC SMART: SM00042; CUB; 2.
CC SMART: SM00192; LDla; 3.
CC SMART: SM00020; TRYPSIN; 1.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS01209; LDLa_1; 2.
CC PROSITE: PS00068; LDLa_2; 4.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.

```

DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Signal-anchor: Glycoprotein; Hydrolase; Serine protease;  
 KW Transmembrane: Repeat.  
 FT DOMAIN 1 55  
 FT TRANSMEM 56 76  
 FT DOMAIN 77 855  
 FT DOMAIN 214 334  
 FT DOMAIN 340 447  
 FT DOMAIN 452 487  
 FT DOMAIN 487 524  
 FT DOMAIN 524 560  
 FT DOMAIN 566 603  
 FT DOMAIN 615 854  
 FT ACT\_SITE 656 656  
 FT ACT\_SITE 711 711  
 FT ACT\_SITE 805 805  
 FT CARBOHYD 109 109  
 FT CARBOHYD 302 302  
 FT CARBOHYD 485 485  
 FT CARBOHYD 772 772  
 FT CONFLICT 327 329  
 FT CONFLICT 381 381  
 FT CONFLICT 674 674  
 FT SEQUENCE 855 AA; 94769 MW; 26143132C01P99C9 CRC64;  
 Query Match 39.3%; Score 42; DB 1; Length 855;  
 Best Local Similarity 50.0%; Pred. No. 61;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 Oy 3 TPOVEVPSQSHI 14  
 1:|||||:  
 Db 367 TWNIEVPNNQHV 378

RESULT 12  
 AROC\_YEAST  
 ID AROC\_YEAST STANDARD; PRT; 376 AA.  
 AC P28777;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Chorismate synthase (EC 4.6.1.4) (5-enolpyruvylshikimate-3-phosphate  
 DE phospholyase).  
 GN ARO2 OR YGL148W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=92114793; PubMed=1837329;  
 RA Jones D.G.L., Reusser U., Braus G.H.;  
 RT "Molecular cloning, characterization and analysis of the regulation  
 RT of the ARO2 gene, encoding chorismate synthase, of Saccharomyces  
 RT cerevisiae.";  
 RL Mol. Microbiol. 5:2143-2152(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1769;  
 RX MEDLINE=97197983; PubMed=9046099;  
 RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
 RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
 RT TPI1, MRF1 genes and six new open reading frames.";  
 RL Yeast 13:177-182(1997).  
 RN [3]  
 RP CATALYTIC ACTIVITY: 5-O-(1-carboxyvinyl)-3-phosphoshikimate =  
 CC chorismate + phosphate.  
 CC -1- COPACOR: REDUCED FLAVIN.  
 CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).

CC -1- INDUCTION: BY AMINO ACID STARVATION.  
 CC -1- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: X60190; CAA42745.1; -  
 CC EMBL: X99960; CAA68214.1; -  
 CC EMBL: 472670; CAA96860.1; -  
 CC PIR: S17246; S17246.  
 CC SGD: S0003116; ARO2.  
 CC InterPro: IPR000453; Chorismate\_synth.  
 CC Pfam: PF01264; Chorismate\_synth.  
 CC ProDom: PD02941; Chorismate\_synth. 1.  
 DR PROSITE: PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
 DR PROSITE: PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
 DR PROSITE: PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
 KW Lyase; Aromatic amino acid biosynthesis.  
 SO SEQUENCE 376 AA; 40838 MW; AF3AF65605B91E8E CRC64;  
 Query Match 38.8%; Score 41.5; DB 1; Length 376;  
 Best Local Similarity 62.5%; Pred. No. 31;  
 Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
 Oy 1 GEFQVQ-VEVPSQSHI 15  
 1:|||||:  
 Db 258 GSGFGQSVPSGSKND 273

RESULT 13  
 GABT\_HUMAN  
 ID GABT\_HUMAN STANDARD; PRT; 500 AA.  
 AC P80404;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 4-aminobutyrate aminotransferase, mitochondrial precursor  
 DE (EC 2.6.1.19) (Gamma-amino-N-butyrate transaminase) (GABA  
 DE transaminase) (GABA aminotransferase) (GABA-AT).  
 GN ABAT OR GABAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95237607; PubMed=7721088;  
 RA Osei Y.D., Churchich J.E.;  
 RT "Screening and sequence determination of a cDNA encoding the human  
 RT brain 4-aminobutyrate aminotransferase.";  
 RL Gene 155:185-187(1995).  
 RN [2]  
 RP SEQUENCE OF 368-465 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Liver;  
 RX MEDLINE=95154329; PubMed=7851425;  
 RA de Biase D., Barra D., Simeco M., John R.A., Bossa F.;  
 RT "Primary structure and tissue distribution of human 4-aminobutyrate  
 RT aminotransferase.";  
 RL Eur. J. Biochem. 227:476-480(1995).  
 RN [3]  
 RP VARIANT IYS-220.  
 RX MEDLINE=99336116; PubMed=10407778;  
 RA Medina-Kauwe L.K., Tobin A.C., De Weirleir L., Jaeken J., Jakobs C.,  
 RA Nyhan W.L., Gibson K.M.;  
 RT "4-aminobutyrate aminotransferase (GABA-transaminase) deficiency.";  
 RL J. Inher. Metab. Dis. 22:414-427(1999).  
 CC -1- CATALYTIC ACTIVITY: 4-aminobutanate + 2-oxoglutarate = succinate

```
CC      semialdehyde + L-glutamate.
CC      -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC      -1- SUBUNIT: MONOMER (PROBABLE).
CC      -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC      -1- TISSUE SPECIFICITY: LIVER > PANCREAS > BRAIN > KIDNEY > HEART >
CC      PLACENTA.
CC      -1- DISEASE: DEFECTS IN ABAT ARE A CAUSE OF GABA-AT DEFICIENCY WHOSE
CC      PHENOTYPE INCLUDES PSYCHOMOTOR RETARDATION, HYPOTONIA,
CC      HYPERREFLEXIA, LETHARGY, REFRACTORY SEIZURES, AND EEG
CC      ABNORMALITIES.
CC      -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC      AMINOTRANSFERASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L32961; AAA74449.1; -.
DR      HSSP; P80147; 1GTX.
DR      MIM; 137150; -.
DR      InterPro: IPR000954; AminoTran_3.
DR      Pfam; PF00202; aminoTran_3; 1.
DR      PROSITE; PS00600; AA_TRANSF_1; 1.
DR      TRANSFASER; AminoTran; Pyridoxal phosphate;
DR      Neurotransmitter degradation; Mitochondrion; Transil peptide;
DR      Disease mutation.
FT      TRANSIT 1 28 MITOCHONDRION.
FT      CHAIN 29 500 4-AMINOBUTYRATE AMINOTRANSFERASE.
FT      BINDING 357 357 PYRIDOXAL PHOSPHATE.
FT      VARIANT 220 220 R -> K (IN GABA-AT DEFICIENCY; 258
FT      REDUCTION IN ACTIVITY).
FT      /FTId=VAR.008883.
FT      CONFLICT 109 109 D -> H (IN REF. 2).
FT      CONFLICT 113 113 E -> L (IN REF. 2).
FT      CONFLICT 132 132 V -> G (IN REF. 2).
FT      CONFLICT 191 191 K -> Q (IN REF. 2).
FT      CONFLICT 204 204 W -> G (IN REF. 2).
FT      CONFLICT 216 216 S -> A (IN REF. 2).
FT      CONFLICT 268 268 G -> R (IN REF. 2).
FT      CONFLICT 320 320 C -> G (IN REF. 2).
FT      CONFLICT 366 366 L -> H (IN REF. 2).
SQ      SEQUENCE 500 AA; 56557 MW; 41199085693F80AD CRC64;
OY      3 TFOVEVPGSOHIDSQKKA 20
Db      17 TYRLVPGSRHI-SQANA 33
Query Match 38.8%; Score 41.5; DB 1; Length 500;
Best Local Similarity 55.6%; Pred. No. 41;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
```

```
RA      Henrich B., Monnerjahn U., Piapp R.;
RT      "Peptidase D gene (pepD) of Escherichia coli K-12: nucleotide
RT      sequence, transcript mapping, and comparison with other peptidase
RT      genes."
RT      J. Bacteriol. 172:4641-4651(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655.
RX      MEDLINE-9742617; PubMed-9278503.
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12."
RT      Science 277:1235-1244(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / W3110.
RA      Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA      Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA      Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT      "Systematic sequencing of the Escherichia coli genome: analysis of the
RT      4.0 - 6.0 min (189,987 - 281,416bp) region."
RT      Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RA      Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA      Davis K., Federle N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA      Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA      Davis R.W.;
RT      Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN      [5]
RP      IDENTIFICATION OF PROTEIN.
RX      MEDLINE-93027135; PubMed-1408743;
RA      Pel H.J., Rep M., Grivell L.A.;
RT      "Sequence comparison of new prokaryotic and mitochondrial members of
RT      the polypeptide chain release factor family predicts a five domain
RT      model for release factor structure."
RT      Nucleic Acids Res. 20:4423-4428(1992).
CC      -1- FUNCTION: COULD BE PROTEIN FACTOR INVOLVED IN TRANSLATIONAL
CC      TERMINATION.
CC      -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
CC      FACTORS FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M34034; -. NOT ANNOTATED_CDS.
DR      EMBL; AE000132; AAC73340.1; -.
DR      EMBL; D83536; BAA77905.1; -.
DR      EMBL; U70214; AAB08656.1; -.
DR      EcoGene; EG11496; prfH.
DR      InterPro: IPR000352; Pep_rel_factor_1.
DR      Pfam; PF00472; RF-1; 1.
DR      PROSITE; PS00745; RF_PROK_1; 1.
KM      Protein biosynthesis, Complete proteome.
FT      DOMAIN 104 107 POLY-GLN.
FT      CONFLICT 1 6 MG1KR -> MLEETGRYSDTLRSAVLSIDGDNAMVLSSES
FT      (IN REF. 2 AND 4).
SQ      SEQUENCE 141 AA; 16177 MW; 9D2EBD9AF7A04831 CRC64;
OY      2 ETFOVEVPGSOHIDSQKKA 21
Db      49 ETLRSGPGGGOHVKNKTSADV 68
Query Match 38.3%; Score 41; DB 1; Length 141;
Best Local Similarity 35.0%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
```



```

RESULT 15
YP68_CAEEL STANDARD: PRT; 313 AA.
ID YP68_CAEEL
AC 009217;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 37.0 kDa protein B0495.8 in chromosome II.
GN B0495.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Kirsten J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO YEAST YDL087C AND S.POMBE SPC16A11.13.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U21317; AAA62527.1; -
DR Wormpep; B0495.8; CE01766.
RW Hypothetical protein.
SQ SEQUENCE 313 AA; 36977 MW; 000D2327621BFED0 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 313;
Best Local Similarity 46.2%; Pred No. 30;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 EYVGSQHIDSQK 19
::||||:|::|
DB 11 QLMGSQHVDNKEX 23

```

Search completed: October 24, 2002, 15:21:04  
 Job time : 9.26229 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:16:03 : Search time 26.1639 Seconds  
(without alignments)  
138.851 Million cell updates/sec

Title: US-09-786-648-5

Perfect score: 107

Sequence: 1 GETFOVEVPGSOHIDSOKKAI 21

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19: \*  
1: sp.archaea: \*  
2: sp.bacteria: \*  
3: sp.fungi: \*  
4: sp.human: \*  
5: sp.invertebrate: \*  
6: sp.mammal: \*  
7: sp.mhc: \*  
8: sp.organelle: \*  
9: sp.phage: \*  
10: sp.plant: \*  
11: sp.prodent: \*  
12: sp.virus: \*  
13: sp.vertebrate: \*  
14: sp.unclassified: \*  
15: sp.virus: \*  
16: sp.bacteriap: \*  
17: sp.archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	94.4	124	2	Q57193 vibrio chol
2	101	94.4	124	2	Q9RP15 vibrio chol
3	101	94.4	124	2	Q56635 vibrio chol
4	101	94.4	124	2	Q93V32 escherichia
5	101	94.4	124	2	Q94M01 vibrio phag
6	95	88.8	103	2	Q9R646 vibrio chol
7	54	50.5	508	4	Q99987 homo sapien
8	52	48.6	552	10	Q9LJ50 arabidopsis
9	49	45.8	255	10	Q9C815 arabidopsis
10	49	45.8	257	10	Q9C875 arabidopsis
11	49	45.8	428	10	Q91GM2 oryza sativ
12	47	43.9	392	10	Q40742 oryza sativ
13	47	43.9	395	5	Q9NKM5 drosophila
14	46	43.0	91	10	Q9SYF4 arabidopsis
15	46	43.0	574	10	Q9SXC9 arabidopsis
16	45.5	42.5	1166	3	Q9PE40 usullago ma

17	45	42.1	439	10	Q9LY67 arabidopsis
18	45	42.1	2732	12	Q9J3F2 murine hepa
19	45	42.1	2733	12	Q9PYA2 murine hepa
20	45	42.1	2733	12	Q9J226 murine hepa
21	45	42.1	2733	12	Q9J3E8 murine hepa
22	44.5	41.6	565	10	Q22511 vitis vinif
23	44	41.1	293	10	Q49876 lupinus alb
24	44	41.1	330	11	Q9QZL7 mus musculu
25	44	41.1	374	5	Q9J3M9 caenorhabd
26	44	41.1	467	6	Q9N136 ovine aries
27	44	41.1	467	11	Q91VDO mus musculu
28	44	41.1	918	3	Q9P606 neurospora
29	43.5	40.7	641	10	Q9SKB2 arabidopsis
30	43.5	40.7	641	10	Q93Z40 arabidopsis
31	43	40.2	356	5	Q9NFP90 leishmania
32	43	40.2	368	10	Q9SN44 arabidopsis
33	43	40.2	843	12	Q9QMN7 hepatitis b
34	43	40.2	843	12	Q91817 hepatitis b
35	43	40.2	849	16	Q98LD6 rhizobium l
36	43	40.2	1714	13	Q90095 gallus galli
37	43	40.2	1610	13	Q90824 gallus galli
38	42	39.3	204	16	Q9HTA0 pseudomonas
39	42	39.3	330	16	Q99VM1 staphylococ
40	42	39.3	427	13	Q42099 cyprinus ca
41	42	39.3	672	16	Q9A818 caulobacter
42	42	39.3	697	5	Q965W6 caenorhabd
43	42	39.3	802	10	Q947W6 oryza sativ
44	42	39.3	945	10	Q94DZ7 oryza sativ
45	42	39.3	946	10	Q94E00 oryza sativ

#### ALIGNMENTS

RESULT 1  
ID 057193 PRELIMINARY: PRT: 124 AA.  
AC 057193;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).  
GN CTXB  
OS vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL STRAIN 569B;  
RX MEDLINE=91355224; PubMed=1683840;  
RA Dams E., De Wolf M., Dierick W.;  
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae  
RT classical strain 569B.";  
RT Biochim. Biophys. Acta 1090:139-141(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL BIOTYPE 569B;  
RA Shi C., Cao C., Zhang J., Ma Q.;  
RL Chin. Biochem. J. 9:395-399(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLASSICAL BIOTYPE 569B;  
RA Xu L.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X58785; CAA41591.1; -;  
DR EMBL; U25679; AAC34728.1; -;  
DR EMBL; A00931; CAA00098.1; -;  
DR HSSP; P01556; 2CHB;  
DR InterPro; IPR001835; Enterotoxin\_B.  
DR Pfam; PF01376; Enterotoxin\_B; 1.  
DR PRINTS; PR00772; ENTEROTOXINB.  
DR ProDom; PD012805; Enterotoxin\_B; 1.  
KW Signal.

```

FT SIGNAL 1 21 POTENTIAL.
FF CHAIN 124 124 CHOLERA TOXIN B PROTEIN (CTB).
SQ SEQUENCE 124 AA; 13919 MW; D6F83FFF792AE3 CRC64;

Query Match
Best Local Similarity 94.4%; Score 101; DB 2; Length 124;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
   | |||||
DB 66 GATFQVEVPGSHIDSOKKAI 86

RESULT 2
ID O9RP15 PRELIMINARY; PRT: 124 AA.
AC O9RP15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHOLERA ENTEROTOXIN B-SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNIH002;
RA Shin H.-J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RT cassette from Vibrio cholerae KNIH002 isolated in Korea.";
RL Misalimurug Holji 35:205-210(1999).
DR EMBL; AF175708; AAD51360.1; -.
DR HSSP; P01556; 2CHB.
DR InterPro; IPR01835; Enterotoxin_B.
DR Pfam; PF01376; Enterotoxin_B; 1.
DR PRINTS; PR00772; ENTEROTOXINB.
DR Prodom; PD012805; ENTEROTOXINB.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FFE793E5B9 CRC64;

Query Match
Best Local Similarity 94.4%; Score 101; DB 2; Length 124;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GETFOVEVPGSHIDSOKKAI 21
   | |||||
DB 66 GATFQVEVPGSHIDSOKKAI 86

RESULT 3
ID O56635 PRELIMINARY; PRT: 124 AA.
AC O56635;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHOLERA TOXIN PRECURSOR.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Miwatani T., Albert M.J.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae
RT strains 854 (O139-Bengal) and S7 (O37) from two outbreaks.";
RL EMBL; D30052; BAA06289.1; -.
DR EMBL; P01556; 2CHB.
DR HSSP; IPR001835; Enterotoxin_B.
DR InterPro; IPR01376; Enterotoxin_B; 1.
DR Pfam; PF01376; Enterotoxin_B; 1.

```

```
DR PRINTS:PR00772; ENTEROTOXINB.  
DR Prodrom: PD012805; Enterotoxin_B; 1.  
KW Signal.  
FT SIGNAL.  
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;  
  
QY 1 GETPQVEVPGSOHIDSOKKAI 21 POTENTIAL.  
| | | | |  
DB 66 GATPQVEVPGSOHIDSOKKAI 86  
  
Query Match 94.4%; Score 101; DB 2; Length 124;  
Best Local Similarity 95.2%; Pred. No. 3.1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.  
  
RESULT 4  
O93V32  
ID O93V32 PRELIMINARY; PRT; 124 AA.  
AC O93V32:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.  
GN LTH B SUBUNIT.  
GN Escherichia coli.  
OC Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1032 (ENTEROTOXIGENIC);  
RA Komase K.; Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN-1032 (ENTEROTOXIGENIC);  
RX MEDLINE=95091056; Pubmed=7998417;  
RA Tanura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,  
RA Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,  
RA Kurata T.;  
RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with  
RT a trace amount of the holotoxin as an adjuvant for nasal influenza  
RT vaccine."  
RL Vaccine 12:1083-1089(1994).  
DR EMBL; AB011677; BAA25726.1;  
SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;  
  
Query Match 94.4%; Score 101; DB 2; Length 124;  
Best Local Similarity 95.2%; Pred. No. 3.1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.  
  
QY 1 GETPQVEVPGSOHIDSOKKAI 21  
| | | | |  
DB 66 GATPQVEVPGSOHIDSOKKAI 86  
  
RESULT 5  
O94M01  
ID O94M01 PRELIMINARY; PRT; 124 AA.  
AC O94M01:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CTXB.  
GN CTXB.  
OS Vibrio phage CTX.  
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
OX NCBI_TaxID=141904;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bhattacharyya T., Nandy R.K., Nair G.B.;  
RT "The entire core region of the ctx-phi (ctx-prophage) in VCE 232, an  
RT environmental strain of V. cholerae.";
```



RA Theologis A., Ecker J.R., Palm C.J., Federpiet N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen Q., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gall J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marzilli A.,  
 RA Maltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utecherack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 DR EMBL: AC051630; AAG51209.1; -  
 DR InterPro: IPR000352; pep\_rel\_factor\_1.  
 DR Pfam: PF00472; RF-1; 1.  
 SQ SEQUENCE 255 AA; 27445 MW; B54F38991FF0D0F8 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 255;  
 Best Local Similarity 45.0%; Pred. No. 3.9;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ETPQVEVPGSGHIDSOKKAI 21  
 Db 101 ETRFVSGGCGHRRKRDASV 120

RESULT 10  
 Q9C875 PRELIMINARY; PRT; 257 AA.  
 AC Q9C875;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DT HYPOTHETICAL 27.6 KDA PROTEIN.  
 GN T1609.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federpiet N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marzilli A.,  
 RA Maltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utecherack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 DR EMBL: AC027035; AAG51290.1; -  
 DR InterPro: IPR000352; pep\_rel\_factor\_1.  
 DR Pfam: PF00472; RF-1; 1.

KW Hypothetical protein.  
 SQ SEQUENCE 257 AA; 27645 MW; 9F85B09C029E97C9 CRC64;  
 Query Match 45.8%; Score 49; DB 10; Length 257;  
 Best Local Similarity 45.0%; Pred. No. 3.9;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ETPQVEVPGSGHIDSOKKAI 21  
 Db 101 ETRFVSGGCGHRRKRDASV 120

RESULT 11  
 Q9LGM2 PRELIMINARY; PRT; 428 AA.  
 AC Q9LGM2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE ESTS AU056822(S20908).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoidae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P041E11.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0433F09.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP002521; BAA96762.1; -  
 DR EMBL: AP002539; BAB08201.1; -  
 DR HSSP: Q06319; IBC.  
 DR InterPro: IPR001552; Acyl-CoA\_dh.  
 DR Pfam: PF00441; Acyl-CoA\_dh\_1.  
 DR Pfam: PF02770; Acyl-CoA\_dh\_M; 1.  
 DR Pfam: PF02771; Acyl-CoA\_dh\_N; 1.  
 DR PROSITE: PS00073; ACYL-CoA\_DH\_2; UNKNOWN 1.  
 SQ SEQUENCE 428 AA; 46132 MW; 8D34E3658A8E6367 CRC64;

Query Match 45.8%; Score 49; DB 10; Length 428;  
 Best Local Similarity 52.6%; Pred. No. 7;  
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TFOYVEVPGSGHIDSOKKAI 21  
 Db 183 TTAATVPGGWHIDGKRWI 201

RESULT 12  
 Q40742 PRELIMINARY; PRT; 392 AA.  
 AC Q40742;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE OSRAD23.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoidae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN=NIPPONBARE;
RX MEDLINE=97369378; PubMed=9225866;
RA Schultz T.F., Quatrano R.S.;
RT "Characterization and expression of a rice RAD23 gene.";
RL Plant Mol. Biol. 34:557-562(1997).
DR EMBL; U63530; AAB65841.1; -.
DR HSSP; P54725; IDVO.
DR InterPro; IPR000449; UBA.
DR InterPro; IPR000626; ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBO; 1.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
SQ SEQUENCE 392 AA; 41754 MW; BDE08574CC7CABC CRC64;

Query Match 43.9%; Score 47; DB 10; Length 392;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 GETFQVEVPGSQHIDSOKKAI 21
Db 10 GSTFQIEVDSAQKADVKKRI 30

RESULT 13
O9NKM5 PRELIMINARY; PRT; 395 AA.
AC O9NKM5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 43.6 KDA PROTEIN.
GN BG:DS01514.3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydridae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Celniker S.E., Abdayani A., Arcalna T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Faltan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Humastli S.R., Karia K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomostan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Setli H., Smit E., Swirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zietan L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003408; AAF4847.1; -.
DR FlyBase; FBgn0028907; BG:DS01514.3.
RW Hypothetical protein.
SQ SEQUENCE 395 AA; 43561 MW; AAF1CC4ADD3DA73 CRC64;

Query Match 43.9%; Score 47; DB 5; Length 395;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 QVEVPGSQHIDSOKKAI 20

```

```

DB 329 RVSPGSHIDADANA 344

RESULT 14
O9SYF4 PRELIMINARY; PRT; 91 AA.
AC O9SYF4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F511.6 PROTEIN.
GN F511.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,
RA Araujo R., Chao O., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F511 sequence.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006577; AAD25769.1; -.
DR HSSP; P02248; UBI1.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 2.
SQ SEQUENCE 91 AA; 10142 MW; E8766823D6450267 CRC64;

Query Match 43.0%; Score 46; DB 10; Length 91;
Best Local Similarity 47.6%; Pred. No. 3; 8;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 GETFQVEVPGSQHIDSOKKAI 21
Db 58 GKTFNLEVKGERIIQYKNNMI 78

RESULT 15
O9SXC9 PRELIMINARY; PRT; 574 AA.
AC O9SXC9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE T17H3.2 PROTEIN (AT1G27520/T17H3_2).
GN T17H3.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Hwang B., Choi E., Chou J., Altafi H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Khan S., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence.";
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,

```

RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005916; AAD45990.1; -;  
DR EMBL; AY056382; AAL08238.1; -;  
DR HSSP; P32906; 1DL2.  
DR InterPro: IPR001382; Glyco\_hydro\_47.  
DR Pfam: PF01532; Glyco\_hydro\_47; 1.  
DR PRINTS; PR007447; GYHRLASE47.  
DR PRODOM; PD003239; Glyco\_hydro\_47; 1.  
SQ SEQUENCE 574 AA; 65707 MW; 4ACC456DE487EA93 CRC64;

Query Match 43.0%; Score 46; DB 10; Length 574;  
Best Local Similarity 58.8%; Pred. No. 31;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 TFQVEVPGSQHIDSQKK 19  
||| |  
Db 19 TFEVVDPPSQHIEVKK 35

Search completed: October 24, 2002, 15:22:30  
Job time : 28.4139 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:20:33 : Search time 47.2787 Seconds  
(without alignments)  
52.114 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCYUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	7	21	US-09-786-648-2
2	39	100.0	12	21	US-09-786-648-3
3	39	100.0	15	11	US-08-732-371-1
4	39	100.0	15	11	US-08-732-371A-1
5	39	100.0	21	21	US-09-786-648-4
6	39	100.0	21	21	US-09-786-648-5
7	39	100.0	93	5	US-08-110-947-10

8	39	100.0	93	5	US-08-110-947A-26	Sequence 26, Appl
9	39	100.0	93	6	US-08-251-121-26	Sequence 26, Appl
10	39	100.0	103	7	US-08-393-334-2	Sequence 2, Appl
11	39	100.0	103	11	US-08-782-832-15	Sequence 15, Appl
12	39	100.0	103	12	US-08-817-906-21	Sequence 21, Appl
13	39	100.0	103	22	US-09-836-433-14	Sequence 14, Appl
14	39	100.0	116	22	US-09-836-433-20	Sequence 20, Appl
15	39	100.0	119	22	US-09-836-433-22	Sequence 22, Appl
16	39	100.0	124	1	PCT-US99-30747-55	Sequence 55, Appl
17	39	100.0	124	1	PCT-US99-30747-57	Sequence 57, Appl
18	39	100.0	124	18	US-09-470-124-55	Sequence 55, Appl
19	39	100.0	124	18	US-09-470-124-57	Sequence 57, Appl
20	39	100.0	138	13	US-08-914-479-2	Sequence 2, Appl
21	39	100.0	138	13	US-08-914-479A-2	Sequence 2, Appl
22	39	100.0	313	21	US-09-756-983-15	Sequence 15, Appl
23	39	100.0	351	21	US-09-756-983-18	Sequence 18, Appl
24	39	100.0	364	21	US-09-756-983-22	Sequence 22, Appl
25	39	100.0	371	11	US-08-150-305A-3	Sequence 3, Appl
26	39	100.0	371	11	US-08-784-218-6	Sequence 6, Appl
27	39	100.0	371	12	US-08-829-026-5	Sequence 5, Appl
28	35	89.7	142	26	US-60-361-742-1547	Sequence 1547, Ap
29	34	87.2	41	20	US-09-688-051-2967	Sequence 2967, Ap
30	34	87.2	84	20	US-09-617-682A-5651	Sequence 5651, Ap
31	34	87.2	124	18	US-09-417-507-29272	Sequence 29272, A
32	34	87.2	304	1	PCT-US01-08631-32869	Sequence 32869, A
33	34	87.2	346	1	PCT-US01-08631-32459	Sequence 32459, A
34	34	87.2	578	1	PCT-US01-08631-32551	Sequence 32551, A
35	34	87.2	847	1	PCT-US01-08631-34035	Sequence 34035, A
36	33	84.6	51	22	US-09-866-066-35	Sequence 35, Appl
37	33	84.6	124	21	US-09-760-234-7	Sequence 7, Appl
38	33	84.6	131	26	US-60-171-481-1502	Sequence 1502, Ap
39	33	84.6	318	1	PCT-US99-17130-318	Sequence 318, App
40	33	84.6	318	18	US-09-489-847-328	Sequence 328, App
41	33	84.6	379	18	US-09-402-532-1	Sequence 1, Appl
42	33	84.6	379	18	US-09-402-532-3	Sequence 3, Appl
43	33	84.6	382	1	PCT-US01-08582-4	Sequence 4, Appl
44	33	84.6	382	1	PCT-US01-08582-4	Sequence 4, Appl
45	33	84.6	382	18	US-09-402-532-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1

US-09-786-648-2

; Sequence 2, Application US/09786648

; GENERAL INFORMATION:

; APPLICANT: Williams, Neil Andrew

; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci

; TITLE OF INVENTION: Adjuvants

; FILE REFERENCE: 7438

; CURRENT APPLICATION NUMBER: US/09/786,648

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: PCT/GB99/02970

; PRIOR FILING DATE: 1999-09-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: MS DOS

; SEQ ID NO 2

; LENGTH: 7

; TYPE: PRT

; ORGANISM: E. coli

; FEATURE:

; LOCATION: 51...57

; OTHER INFORMATION: Isolated or synthetic EtxB beta2-alpha2 loop fragment derivabl

; OTHER INFORMATION: human variant E. coli

US-09-786-648-2

Query Match 100.0%; Score 39; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3.2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;

Db 1 EVPGSQH 7

## RESULT 2

US-09-786-648-3

Sequence 3, Application US/09786648  
GENERAL INFORMATION:  
APPLICANT: Williams, Neil Andrew  
TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine  
FILE REFERENCE: 7438  
CURRENT APPLICATION NUMBER: US/09/786,648  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: PCT/CB99/02970  
PRIOR FILING DATE: 1999-09-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: MS DOS  
SEQ ID NO 3  
LENGTH: 12  
TYPE: PRT  
ORGANISM: E. coli  
FEATURE:  
LOCATION: 50..61  
OTHER INFORMATION: isolated or synthetic EtxB beta4-alpha2 loop fragment derivable  
US-09-786-648-3

## Query Match

Best Local Similarity 100.0%; Score 39; DB 21; Length 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 2 EVPGSQH 8

## RESULT 3

US-08-732-371-1

Sequence 1, Application US/08732371  
GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371-1

## Query Match

Best Local Similarity 100.0%; Score 39; DB 11; Length 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 2 EVPGSQH 8

## RESULT 4

US-08-732-371A-1

Sequence 1, Application US/08732371A  
GENERAL INFORMATION:  
APPLICANT: MIRELMAN, David  
APPLICANT: MARKS, Robert S.  
TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,371A  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 109519  
FILING DATE: 03-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: MIRELMAN-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-732-371A-1

## Query Match

Best Local Similarity 100.0%; Score 39; DB 11; Length 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 2 EVPGSQH 8

## RESULT 5

US-09-786-648-4

```
; Sequence 4, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 4
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic EcxB beta4-alpha2 loop fragment derivable f
; US-09-786-648-4

Query Match          100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 EVPGSOH 7
        |||
        7 EVPGSOH 13

RESULT 6
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; APPLICANT: Hirst, Timothy Raymond
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic EcxB beta4-alpha2 loop fragment derivable f
; US-09-786-648-5

Query Match          100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 EVPGSOH 7
        |||
        7 EVPGSOH 13

RESULT 7
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COCKLE, Stephen A
```

```
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Hwy.
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-110-947-10

Query Match          100.0%; Score 39; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 EVPGSOH 7
        |||
        41 EVPGSOH 47

RESULT 8
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; STREET: Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
```

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLOW, Charles W  
REGISTRATION NUMBER: 28, 946  
REFERENCE/DOCKET NUMBER: 1038-303 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
TELEX: 89-9456 LOKPAT  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-110-947A-26

Query Match 100.0%; Score 39; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7  
|||||||  
DB 41 EYVGSQH 47

RESULT 9  
US-08-251-121-26  
Sequence 26, Application US/08251121

GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/251,121  
FILING DATE: 31-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-335  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-251-121-26

Query Match 100.0%; Score 39; DB 6; Length 93;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7  
|||||||  
DB 41 EYVGSQH 47

RESULT 10  
US-08-393-334-2

Sequence 2, Application US/08393334

GENERAL INFORMATION:  
APPLICANT: Loomore, Sheena M.  
APPLICANT: YACOUB, Reza K.  
APPLICANT: zealey, Gavin R.  
APPLICANT: Klein, Michel H.  
TITLE OF INVENTION: Expression of Gene Products from  
Genetically Manipulated Strains of Bordetella  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 330 University Avenue, Suite 701  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,334  
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-417  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-595-1155  
TELEFAX: 416-595-1163  
TELEX: 065-24567 Simbas  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-393-334-2

Query Match 100.0%; Score 39; DB 7; Length 103;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7  
|||||||  
DB 51 EYVGSQH 57

RESULT 11  
US-08-782-832-15

Sequence 15, Application US/08782832

GENERAL INFORMATION:  
APPLICANT: Aintzen, Charles J.  
APPLICANT: Mason, Hugh S.  
APPLICANT: Haq, Tariq A.  
TITLE OF INVENTION: PRODUCTION OF AN ORALLY IMMUNOGENIC  
BACTERIAL PROTEIN IN TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewlett, Kimball & Krieger

STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/782,832  
FILING DATE: 13-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/328,716  
FILING DATE: 24-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kettelberger, Denise M.  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 36170/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-782-832-15

Query Match 100.0%; Score 39; DB 11; Length 103;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSGSH 7  
|||||||  
Db 51 EYPSGSH 57

RESULT 12  
US-08-817-906-21  
Sequence 21, Application us/08817906  
GENERAL INFORMATION:  
APPLICANT: Charles J. Arttzen, Hugh S. Mason, John D. Clements,  
APPLICANT: and Farid A. Haq  
TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,906  
FILING DATE: 08/04/97  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: 36170/3P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151

TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-817-906-21

Query Match 100.0%; Score 39; DB 12; Length 103;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSGSH 7  
|||||||  
Db 51 EYPSGSH 57

RESULT 13  
US-09-836-433-14  
Sequence 14, Application US/09836433  
GENERAL INFORMATION:  
APPLICANT: Yuki, Yoshikazu  
APPLICANT: Uda, Shigeo  
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
FILE REFERENCE: Not Assigned  
CURRENT APPLICATION NUMBER: US/09/836,433  
CURRENT FILING DATE: 2001-04-16  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 14  
LENGTH: 103  
TYPE: PRT  
ORGANISM: Vibrio Cholerae  
US-09-836-433-14

Query Match 100.0%; Score 39; DB 22; Length 103;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSGSH 7  
|||||||  
Db 51 EYPSGSH 57

RESULT 14  
US-09-836-433-20  
Sequence 20, Application US/09836433  
GENERAL INFORMATION:  
APPLICANT: Yuki, Yoshikazu  
APPLICANT: Uda, Shigeo  
TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE  
FILE REFERENCE: Not Assigned  
CURRENT APPLICATION NUMBER: US/09/836,433  
CURRENT FILING DATE: 2001-04-16  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 20  
LENGTH: 116  
TYPE: PRT  
ORGANISM: synthetic construct  
US-09-836-433-20

Query Match 100.0%; Score 39; DB 22; Length 116;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSGSH 7  
|||||||  
Db 51 EYPSGSH 57

RESULT 15  
US-09-836-433-22

```
; Sequence 22, Application US/09836433
; GENERAL INFORMATION:
; APPLICANT: Yuki, Yoshikazu
; APPLICANT: Udaoka, Shigezo
; TITLE OF INVENTION: HYBRID PROTEINS FOR AUTOIMMUNE DISEASE
; FILE REFERENCE: Not Assigned
; CURRENT APPLICATION NUMBER: US/09/836,433
; CURRENT FILING DATE: 2001-04-16
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 119
; TYPE: prt
; ORGANISM: synthetic construct
; US-09-836-433-22
```

```
Query Match      100.0%; Score 39; DB 22; Length 119;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 EYVGSQH 7
         |||||
Db      51 EYVGSQH 57
```

Search completed: October 24, 2002, 15:30:56  
Job time : 49.2787 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:22:33 ; Search time 10.9016 Seconds  
(without alignments)  
71.321 Million cell updates/sec

Title:	US-09-786-6
Perfect score:	39
Sequence:	1 EVPGSQH 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 469622
```

```
Minimum DB seq length: 0
Maximum DB seq length: 100
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : A\_Geneseq\_032802:\*

1	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1980.DAT *
2	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1981.DAT *
3	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1982.DAT *
4	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1983.DAT *
5	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1984.DAT *
6	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1985.DAT *
7	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1986.DAT *
8	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1987.DAT *
9	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1988.DAT *
10	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1989.DAT *
11	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1990.DAT *
12	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1991.DAT *
13	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1992.DAT *
14	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1993.DAT *
15	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1994.DAT *
16	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1995.DAT *
17	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1996.DAT *
18	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1997.DAT *
19	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1998.DAT *
20	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA1999.DAT *
21	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA2000.DAT *
22	/SIDS1/gcgdata/geneseq/geneseqp_emb1/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	7	21	AA87460	Cholera toxin B/er
2	39	100.0	8	21	AA87464	Cholera toxin B/er
3	39	100.0	12	21	AA87461	Cholera toxin B/er
4	39	100.0	15	10	AA87498	CTP3 epitope of tlo
5	39	100.0	15	16	AA85125	Cholera toxin B ar
6	39	100.0	21	21	AA87462	Cholera toxin B/er
7	39	100.0	21	21	AA87463	E. coli heat labil
8	39	100.0	23	16	AA87648	Residues 50-64 of
9	39	100.0	26	4	AA80265	Sequence of amino
10	39	100.0	41	6	AA85043	Network polymer wh
11	39	100.0	46	6	AA850436	Network polymer wh

12	39	100.0	47	4	AAP30600.
13	39	100.0	93	16	AAR72545.
14	39	100.0	93	20	AAV41810.
15	39	100.0	93	20	AAW952222.
16	39	100.0	93	21	AAV683656.
17	39	100.0	93	22	AAAG623333.
18	34	87.2	51	21	AAAG263333.
19	33	84.6	15	22	AAB743222.
20	31	79.5	52	22	AAU555222.
21	30	76.9	76	22	AAO099555.
22	30	76.9	79	20	AAV128919.
23	30	76.9	84	22	AAAB46666.
24	30	76.9	92	21	AAO006333.
25	30	76.9	93	20	AAV117555.
26	30	76.9	95	22	ABBI79666.
27	29	74.4	26	20	AAV278222.
28	29	74.4	29	22	ABAB8719.
29	29	74.4	29	22	ABB338989.
30	29	74.4	29	22	ABBI93388.
31	29	74.4	29	22	AAAB46666.
32	29	74.4	29	22	AAAB70666.
33	29	74.4	29	22	AAAM49303.
34	29	74.4	29	22	AAAB73555.
35	29	74.4	29	22	AAW026555.
36	29	74.4	48	22	AAOI23555.
37	29	74.4	83	22	AAAB91444.
38	29	74.4	86	20	AAV079555.
39	29	74.4	86	20	AAV078555.
40	28	71.8	33	22	AAAG76915.
41	28	71.8	33	22	AAAB31655.
42	28	71.8	53	22	AAAB33955.
43	28	71.8	58	20	AAV059666.
44	28	71.8	58	21	AAV088666.
45	28	71.8	65	22	AAOI06444.

## ALIGNMENTS

RESULT 1	
AA87460	
ID	AA87460 standard; peptide; 7 AA.
XX	
AC	AA87460;
XX	
DE	03-JUL-2000 (first entry)
XX	
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:2.
XX	
KW	Cholera toxin B subunit B; CTxB; heat labile enterotoxin subunit B; EtxB
KW	beta-2-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
KW	adjuvant; immune disorder; diarrhoea.
XX	
OS	Vibrio cholerae.
OS	Escherichia coli.
XX	
PN	W0200014114-A1.
XX	
PD	16-MAR-2000.
XX	
PF	07-SEP-1999; 99WO-GB02970.
XX	
PR	07-SEP-1998; 98GB-0019484.
XX	
PPA	(UYBR-) UNIV BRISTOL.
XX	
PI	Williams NA, Hirst TR;
XX	
DR	WPI, 2000-256943/22.
XX	
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as
PT	immunomodulators and for treating diarrhea and which do not bind the
PT	glycolipid receptor GM-1 -

XX Claim 1, Page 13, 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat

CC labile enterotoxin (Etx) and its closely related homologue, cholera

CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous

CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

CC composed of one A subunit and five identical B subunits. The A subunit

CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)

CC facilitate the entry of subunit A into the host cell via the binding and

CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible

CC for some of the effects of Etx and Ctx, it has been found that certain

CC effects of the toxins, such as immunomodulation, are not mediated

CC through GM-1 binding. The peptides of the invention are fragments of the

CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as

CC normal EtxB and CtxB subunits, except that they do not bind or cross link

CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.

CC Therefore, the peptides may be used in the production of a composition

CC for treating, preventing and/or modulating a disease associated with an

CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463

CC represent preferred peptides of the invention, AAY87460 being

CC particularly preferred.

SQ Sequence 7 AA:

Query Match 100.0%; Score 39; DB 21; Length 7;

Best Local Similarity 100.0%; Pred. No. 6.4e+05; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
| | | | | | |  
DB 1 EVPGSOH 7

RESULT 2  
AAY87464 standard; peptide; 8 AA.

AC AAY87464;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived immunomodulatory peptide.

KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KW adjuvant; immune disorder; diarrhoea.

OS Vibrio cholerae.

OS Escherichia coli.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GB02970.

PR 07-SEP-1998; 98GB-0019484.

PA (UYBR-) UNITV BRISTOL.

PI Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as

PT immunomodulators and for treating diarrhoea and which do not bind the

PT glycolipid receptor GM-1 -

PS Example 5; Page 45; 62pp; English.

CC The invention relates to peptide fragments of the Escherichia coli heat

CC labile enterotoxin (Etx) and its closely related homologue, cholera

CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous

CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are

CC composed of one A subunit and five identical B subunits. The A subunit

CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-

CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)

CC facilitate the entry of subunit A into the host cell via the binding and

CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible

CC for some of the effects of Etx and Ctx, it has been found that certain

CC effects of the toxins, such as immunomodulation, are not mediated

CC through GM-1 binding. The peptides of the invention (AAY87461-Y87463) are

CC fragments of the beta-4-alpha-2 loop of EtxB and/or CtxB, exerting the

CC same effects as normal EtxB and CtxB subunits, except that they do not

CC bind or cross link GM-1. They may be used in medicine as an

CC immunomodulator or adjuvant. They may also be used as an inhibitor for

CC toxin-induced diarrhoea. Therefore, the peptides may be used in the

CC production of a composition for treating, preventing and/or modulating a

CC disease associated with an immune disorder and/or toxin-induced

CC diarrhoea. Sequences AAY87464-Y87465 represent peptides used in an

CC exemplification of the present invention to assess whether a peptide

CC corresponding to a portion of the beta-4-alpha-2 loop of EtxB and CtxB

CC has immunomodulatory effects. Peptide AAY87464 corresponds to residues

CC 51-58 of the EtxB/CtxB beta-4-alpha-2 loop, and peptide AAY87465 is a

CC randomly selected control peptide.

SQ Sequence 8 AA:

Query Match 100.0%; Score 39; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 6.4e+05; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
| | | | | | |  
DB 1 EVPGSOH 7

RESULT 3  
AAY87461 standard; peptide; 12 AA.

AC AAY87461;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.

KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KW adjuvant; immune disorder; diarrhoea.

OS Vibrio cholerae.

OS Escherichia coli.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GB02970.

PR 07-SEP-1998; 98GB-0019484.

PA (UYBR-) UNITV BRISTOL.

PI Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

XX Derivatives of Escherichia coli heat labile enterotoxins useful as

PT immunomodulators and for treating diarrhoea and which do not bind the

PT glycolipid receptor GM-1 -

PS Disclosure; Page 15; 62pp; English.



```

XX CC The invention relates to peptide fragments of the Escherichia coli heat
CC CC labile enterotoxin (Etx) and its closely related homologue, cholera
CC CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC CC composed of one A subunit and five identical B subunits. The A subunit
CC CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC CC facilitate the entry of subunit A into the host cell via the binding and
CC CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC CC for some of the effects of Etx and Ctx, it has been found that certain
CC CC effects of the toxins, such as immunomodulation, are not mediated
CC CC through GM-1 binding. The peptides of the invention are fragments of the
CC CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC CC normal EtxB and CtxB subunits, except that they do not bind or cross link
CC CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC CC They may also be used as an inhibitor for toxin-induced diarrhoea.
CC CC Therefore, the peptides may be used in the production of a composition
CC CC for treating, preventing and/or modulating a disease associated with an
CC CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC CC represent preferred peptides of the invention, AAY87460 being
CC CC particularly preferred.
XX SQ Sequence 12 AA:
XX
XX Query Match 100.0%; Score 39; DB 21; Length 12;
XX Best Local Similarity 100.0%; Pred. No. 0.29;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPGSOH 7
XX |||||
XX Db 2 EYPGSOH 8
XX
XX RESULT 4
XX AAP93498
XX ID AAP93498 standard; protein: 15 AA.
XX
XX AC AAP93498;
XX
XX DP 03-MAY-1990 (first entry)
XX
XX DE CTP3 epitope of the Cholera toxin B subunit.
XX
XX KW CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;
XX KW vaccine; immunotherapy; ds;
XX PN *W089A0967EA.
XX
XX PD 16-NOV-1989.
XX
XX PF 05-MAY-1989; 89WO-US01932.
XX
XX PR 05-MAY-1988; 88US-0190570.
XX
XX PA (PRAAX-) PRAAXS BIOLOGICS INC.
XX PA (STRD ) LEYLAND STANDFORD JUNIOR UNIV.
XX
XX PI Marjarian WR, Stocker BAD, Newton SMK;
XX
XX DR WPI: 1989-356496/48.
XX DR N-PSDB: AAN92414.
XX
XX FT New recombinant flagellin gene including sequence - for heterologous
XX FT epitope, and expressed fusion proteins, useful in vaccines and for prodn.
XX FT of antibodies.
XX
XX PS Disclosure: fig.4B; 137pp; English.
XX
XX CC This sequence corresponds to the CTP3 epitope of the Cholera toxin B
XX CC subunit. The DNA sequence encoding this ligates to othersynthetic
XX CC oligonucleotides to form a new recombinant gene. This encodes
XX CC a flagellin fusion protein which can be used in vaccines for immuno-

```

```

CC CC therapy.
XX SQ Sequence 15 AA:
XX
XX Query Match 100.0%; Score 39; DB 10; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 0.37;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPGSOH 7
XX |||||
XX Db 2 EYPGSOH 8
XX
XX RESULT 5
XX ID AAR85125 standard; peptide: 15 AA.
XX
XX AC AAR85125;
XX
XX DP 13-JUN-1996 (first entry)
XX
XX DE Cholera toxin B antigenic peptide fragment CTP3.
XX
XX KW Conjugate; cholera; B toxin; peptide fragment; microparticulate;
XX KW inert carrier; modified silica; thyroglobulin; oral vaccine;
XX KW immunisation; infection; insoluble; digestive tract; antigen;
XX KW intestines; antibodies; secretory; Iga class.
XX
XX OS Vibrio cholerae.
XX
XX PN W09529701-A1.
XX
XX PD 09-NOV-1995.
XX
XX PF 02-MAY-1995; 95WO-EP01661.
XX
XX PR 03-MAY-1994; 94IL-0109519.
XX
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX
XX PI Marks RS, Mirelman D, Sela M;
XX
XX DR WPI: 1995-403805/51.
XX
XX FT Vaccines for oral immunisation against infecting agents, e.g.
XX FT cholera - comprise a conjugate of an antigen of an infecting agent
XX FT covalently bound to micro:particulate inert carrier, e.g. modified
XX FT aldehyde silica
XX
XX PS Claim 7; Page 25; 40pp; English.
XX
XX CC A compsn. comprising a conjugate of an antigenic cholera B toxin
XX CC peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently
XX CC bound to a microparticulate inert carrier (e.g. modified silica or
XX CC thyroglobulin) can be used as an oral vaccine for immunisation
XX CC against cholera infection. The inert carrier is insoluble in the
XX CC digestive tract, allowing presentation of the antigen in the
XX CC intestines, where it will elicit antibodies mainly of the
XX CC secretory Iga class.
XX
XX SQ Sequence 15 AA:
XX
XX Query Match 100.0%; Score 39; DB 16; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 0.37;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPGSOH 7
XX |||||
XX Db 2 EYPGSOH 8
XX
XX RESULT 6
XX AAY87462

```

ID	AAV87462	standard; peptide; 21 AA.
AC	AAV87462;	-
XX	03-JUL-2000	(first entry)
DE	Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.	
XX	Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;	
XX	beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;	
KW	adjuvant; immune disorder; diarrhoea.	
XX	Vibrio cholerae.	
OS	Escherichia coli.	
XX	WO200014114-A1.	
PD	16-MAR-2000.	
XX	07-SEP-1999;	99WO-GB02970.
PF	07-SEP-1998;	98GB-0019484.
XX	(UYBR-) UNITV BRISTOL.	
PA	Williams NA, Hirst TR;	
PL	WPI: 2000-256943/22.	
DR	Derivatives of Escherichia coli heat labile enterotoxins useful as	
XX	immunomodulators and for treating diarrhoea and which do not bind the	
PT	glycolipid receptor GM-1 -	
XX	Disclosure; Page 15; 62pp; English.	
PS	The invention relates to peptide fragments of the Escherichia coli heat	
CC	labile enterotoxin (Etx) and its closely related homologue, cholera	
CC	toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous	
CC	GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are	
CC	composed of one A subunit and five identical B subunits. The A subunit	
CC	is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-	
CC	ribosyltransferase activity, while the B subunits (EtxB and CtxB)	
CC	facilitate the entry of subunit A into the host cell via the binding and	
CC	cross-linking of GM-1 receptors. Although GM-1 binding is responsible	
CC	for some of the effects of Etx and Ctx, it has been found that certain	
CC	effects of the toxins, such as immunomodulation, are not mediated	
CC	through GM-1 binding. The peptides of the invention are fragments of the	
CC	beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as	
CC	normal EtxB and CtxB subunits, except that they do not bind or cross link	
CC	GM-1. They may be used in medicine as an immunomodulator or adjuvant.	
CC	They may also be used as an inhibitor for toxin-induced diarrhoea.	
CC	Therefore, the peptides may be used in the production of a composition	
CC	for treating, preventing and/or modulating a disease associated with an	
CC	immune disorder and/or toxin-induced diarrhoea. Sequences AAV87460-Y87463	
CC	represent preferred peptides of the invention, AAV87460 being	
CC	particularly preferred.	
XX	Sequence	21 AA;
XX		
QY	Query Match	100.0%; Score 39; DB 21; Length 21;
	Best Local Similarity	100.0%; Pred. No. 0.52;
Matches	7; Conservative	0; Mismatches 0; Indels 0; Gaps 0
DB	1 EYPGSOH 7	
	7 EYPGSOH 13	
RESULT 7		
AAV87463		
ID	AAV87463	standard; peptide; 21 AA.
XX	AAV87463;	
CC		

XX	03-JUL-2000	(first entry)
DT		
XX	E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.	
DE		
XX	Heat labile enterotoxin subunit B; EtxB;	
XX	beta-M-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;	
KW	adjuvant; immune disorder; diarrhoea.	
KW		
OS	Escherichia coli.	
XX		
XX	WO200014114-A1.	
PN		
PI	16-MAR-2000.	
XX		
XX	07-SEP-1999;	99WO-GB02970.
PF		
PR	07-SEP-1998;	98GB-0019484.
XX		
PA	(UYBR-) UNIV BRISTOL.	
XX		
XX	Williams NA, Hirst TR;	
PI		
DR	WPI: 2000-256943/22.	
XX		
PT	Derivatives of Escherichia coli heat labile enterotoxins useful as	
PT	immunomodulators and for treating diarrhoea and which do not bind the	
PT	glycolipid receptor GM-1.	
XX		
PS	Disclosure; Page 15; 62pp; English.	
XX		
XX	The invention relates to peptide fragments of the Escherichia coli heat	
CC	labile enterotoxin (Etx) and its closely related homologue, cholera	
CC	toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous	
CC	GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are	
CC	composed of one A subunit and five identical B subunits. The A subunit	
CC	is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-	
CC	ribosyltransferase activity, while the B subunits (EtxB and CtxB)	
CC	facilitate the entry of subunit A into the host cell via the binding and	
CC	cross-linking of GM-1 receptors. Although GM-1 binding is responsible	
CC	for some of the effects of Etx and Ctx, it has been found that certain	
CC	effects of the toxins, such as immunomodulation, are not mediated	
CC	through GM-1 binding. The peptides of the invention are fragments of the	
CC	beta-1-alpha-2 loop of EtxB and/or CtxB, exerting the same effects as	
CC	normal EtxB and CtxB subunits, except that they do not bind or cross link	
CC	GM-1. They may be used in medicine as an immunomodulator or adjuvant.	
CC	Therefore, the peptides may be used in the production of a composition	
CC	for treating, preventing and/or modulating a disease associated with an	
CC	immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-B	
CC	represent preferred peptides of the invention, AAY87460 being	
CC	particularly preferred.	
XX		
XX	Sequence	21 AA;
XX		
XX	Query Match	100.0%; Score 39; DB 21; Length 21;
XX	Best Local Similarity	100.0%; Pred. No. 0.52;
XX	Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 EVPSGQH 7	
XX		
XX	7 EVPSGSH 13	
Db		
XX		
XX	RESULT 8	
XX	AAR76748	
ID	AAR76748 standard; Protein; 23 AA.	
XX	AAR76748;	
XX		
DT	18-MAR-1996	(first entry)
XX		
DE	Residues 50-64 of cholera toxin B subunit and P1mh 224-226.	

```

XX FimH; type 1 fimbriae; organelle: adhesin; alpha-D-mannoside residue;
KM FimA; FimA; receptor binding site; PCR; amplify; ss.
XX Chimeric - Vibrio cholerae.
OS Chimeric - Escherichia coli.
XX
FH Key
FT Misc-difference 1 Location/Qualifiers
FT Peptide 2..4 /note- "Represents FimH residue 224"
FT Peptide 5..19 /note- "Linker peptide"
FT Peptide 20..22 /note- "Cholera toxin B subunit 50-64"
FT Misc-difference 23 /note- "Linker peptide"
FT /note- "Represents FimH residue 226"
XX
PN MO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95MO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX WPI: 1995-275442/36.
XX
DR N-PSDB; AA093061.
XX
PT Receptor specific bacterial adhesins - useful for targeting active
XX compounds and microbial cells to locations of receptors
XX
PS Example 3; Page 58; 152pp; English.
XX
CC This sequence is encoded by a fragment of the the plasmid pLPA93
CC which was used in the production of fimb fusion genes comprising
CC the cholera toxin B subunit inserted into the fimb gene. This insert
CC shows the inclusion of the B subunit into the fimb protein at position
CC 224-226. The chimeric genes were then opt. further modified by insertion
CC of the hepatitis B virus surface antigen pre-S2 region into a different
CC position of the fimb adhesin of type 1 fimbriae. Restriction site handles
CC (BglII-sites) were introduced into the fimb gene, and the foreign
CC epitopes are then inserted in-frame. In the selected positions the
CC insertion of the epitopes did not significantly alter the adhesive
CC function of the fimb protein. The expression of the chimeric proteins
CC on the surface of fimbriae on bacterial hosts illustrated the possibility
CC of using bacterial adhesins as general presenters of foreign antigens and
CC epitopes. These chimeric genes may be used in the production of variant
CC fimb adhesins which may be useful for targeting active compounds
CC and microbial cells to locations comprising selected receptors to which
CC the adhesins bind.
XX
SQ Sequence 23 AA;
XX
Query Match 100.0%; Score 39; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 EYPSGQH 7
Db 6 EYPSGQH 12
XX
RESULT 9
AAP30265
ID AAP30265 standard; Protein; 26 AA.
XX
AC AAP30265;

```

```

XX
DT 21-APR-1992 (first entry)
XX
DE Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
DE carries an Arg at posns. 67 and 73.
XX
KM Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
XX
OS Vibrio cholerae.
XX
PN EP95426-A.
XX
PD 30-NOV-1983.
XX
PF 26-MAY-1983; 83EP-0401052.
XX
PR 26-MAY-1982; 82FR-0009167.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
PA (INSP ) INST PASTEUR.
XX
PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
XX Guyongruaz A, Delmas A;
XX WPI: 1983-834645/49.
XX
DR
XX
PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and
PT medicaments - effective against Escherichia coli and Vibrio
PT cholerae infections, are prepd. by solid phase peptide synthesis
XX
PS Claim 7; Page 11; 13pp; French.
XX
CC The inventors claim cholera toxin B1 subunit sequences which carry
CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing
CC these toxins to cell walls. The peptides are used in the treatment of,
CC and vaccination against, cholera infections and animal and human
CC infections due to E. coli (enterotoxin LT). The medicament may be
CC administered by oral, intraperitoneal, sub-cutaneous or intravenous
CC routes. For vaccines, pref. peptides having 15-30 AAs are injected
CC without carriers. Unit dose when used as a medicament is 50-500mg as
CC a vaccine 1-10mg of active cpd.
XX
SQ Sequence 26 AA;
XX
Query Match 100.0%; Score 39; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 EYPSGQH 7
Db 2 EYPSGQH 8
XX
RESULT 10
AAP50439
ID AAP50439 standard; protein; 41 AA.
XX
AC AAP50439;
XX
DT 01-JAN-1980 (first entry)
XX
DE Network polymer which comprises a series of composite E. coli heat-
DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
XX
KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
XX
OS Synthetic.
XX
PN WO8502611-A.
XX
PD 20-JUN-1985.
XX

```

PF 12-DEC-1984; 84WO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 XX useful for vaccination of mammals against the enterotoxin(s)  
 PS  
 XX Claim 8; Page 100; 120pp; English.  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 41 AA;  
 Query Match 100.0%; Score 39; DB 6; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EYPGSOH 7  
 DB 15 EYPGSOH 21  
 RESULT 11  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX  
 AC AAP50436;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX  
 KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO8502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PE 12-DEC-1984; 84WO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI; 1985-159230/26.  
 XX  
 PT New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 XX useful for vaccination of mammals against the enterotoxin(s)  
 PS  
 XX Claim 8; Page 100; 120pp; English.  
 CC The repeating units are bonded together by intramolecular  
 CC interpolypeptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or

CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 46 AA;  
 Query Match 100.0%; Score 39; DB 6; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EYPGSOH 7  
 DB 17 EYPGSOH 23  
 RESULT 12  
 AAP30600  
 ID AAP30600 standard; protein; 47 AA.  
 XX  
 AC AAP30600;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posn. 35, 67 and 73.  
 XX  
 KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PE 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82EP-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivallie P, Siffert O, Dodin A;  
 XX  
 PI Guyongruaz A, Delmas A;  
 XX  
 DR WPI; 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 8; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posn. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against, cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX  
 SQ Sequence 47 AA;  
 Query Match 100.0%; Score 39; DB 4; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EYPGSOH 7  
 DB 23 EYPGSOH 29

```

RESULT 13
AAR72545
ID AAR72545 standard; peptide; 93 AA.
XX
AC AAR72545;
XX
DT 28-NOV-1995 (first entry)
XX
DE ADP-ribosylating toxin (verotoxin-1 B-subunit).
XX
KW ADP-ribosylating toxin; pertussis holotoxin; B-subunit;
KM active site; E. coli heat labile toxin; verotoxin-1;
XX Bordetella pertussis vaccines.
OS Bacteria sp.
XX
PN EP646599-A.
XX
PD 05-APR-1995.
XX
PF 23-AUG-1994; 94EP-0306219.
XX
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (UYAL-) UNIV ALBERTA.
XX
PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI Oomen R, Read RJ, Stein PE;
XX
DR WPI; 1995-132623/18.
XX
PT New modified forms of pertussis holotoxin - developed using
PT crystalline forms of pertussis holotoxin and its complexes with
PT other molecules
XX
PS Disclosure; Fig 5; 54pp; English.
XX
CC AAR75540-R75545 are structurally equivalent B-subunits from three
CC ADP-ribosylating toxins, pertussis holotoxin (PT), E. coli heat
CC labile toxin (LT), and verotoxin-1 (VT). The structural
CC information obtd. from these comparisons was used to identify
CC sites which contribute to PT's biological activity. By modifying
CC these sites the claimed PT mutants of the invention were produced,
CC they can be used in the development of vaccines against Bordetella
CC pertussis infection.
XX
SQ Sequence 93 AA;
XX
Query Match 100.0%; Score 39; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPSGSH 7
DB 41 EYPSGSH 47
XXXXXXXXXX
RESULT 14
AAV41816
ID AAV41816 standard; peptide; 93 AA.
XX
AC AAV41816;
XX
DT 08-DEC-1999 (first entry)
XX
DE Escherichia coli verotoxin-1 B-subunit.
XX
KW ADP-ribosylating toxin; PT; crystalline pertussis holotoxin; ETA; DT;
KM three-dimensional structure; LT; immunoprotective; infection.
XX
OS Escherichia coli.
XX

```

```

XX
PN US5965385-A.
XX
PD 12-OCT-1999.
XX
PE 06-JUN-1995; 95US-0467974.
XX
PR 22-AUG-1994; 94US-0292968.
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (UYAL-) UNIV ALBERTA.
XX
PI Read RJ, Stein PE, Cockle SA, Loosmore S, Klein MH, Armstrong GD;
PI Hazes B, Oomen RP;
XX
DR WPI; 1999-579908/49.
XX
PT New method for producing modified pertussis holotoxin -
XX
PS Example 3; Fig 5; 41pp; English.
XX
CC A method has been developed of producing a modified pertussis holotoxin,
CC involving analysis of the 3-dimensional form of the crystalline
CC holotoxin. The pertussis holotoxin modification process comprises:
CC (1) identification of at least one amino acid (aa) residue of the
CC holotoxin for modification by analysing the 3-dimensional form of the
CC crystalline holotoxin, in relation to known information of the protein
CC structure and function; (2) effecting mutagenesis (by removing or
CC replacing a nucleotide sequence encoding at least one (aa) of a tox
CC operon; and (3) expressing mutant tox box in a Bordetella organism to
CC produce the modified holotoxin. This method is used for modifying
CC pertussis holotoxin, by studying its 3-dimensional crystalline
CC structure. Modifying the holotoxin, alters its biological properties.
CC By analysing the 3-dimensional crystalline structure of the pertussis
CC holotoxin, functional (aa) which affect biological properties of the
CC pertussis holotoxin can be identified. This can be used to predict (aa)
CC which contribute to the toxicity of the holotoxin to produce
CC immunoprotective, genetically-detoxified analogues of pertussis
CC holotoxin. The present sequence represents an ADP-ribosylating toxin
CC B-subunit peptide used in the exemplification of the present
CC invention.
XX
SQ Sequence 93 AA;
XX
Query Match 100.0%; Score 39; DB 20; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPSGSH 7
DB 41 EYPSGSH 47
XXXXXXXXXX
RESULT 15
AAW95226
ID AAW95226 standard; peptide; 93 AA.
XX
AC AAW95226;
XX
DT 16-MAR-1999 (first entry)
XX
DE E. coli heat-labile toxin (LT) beta-subunit sequence.
XX
KW Pertussis holotoxin; PT; modified; effector; toxicity; cell binding;
KW enzyme; ADP-ribosylation; toxin; crystalline; X-ray crystallography;
KM structural analysis; interacting site; mitogenicity; adjuvantivity;
KM heat-labile; LT.
XX
OS Escherichia coli.
XX
PN US5856122-A.
XX

```

```

XX 05-JAN-1999.
PD
XX
XX 22-AUG-1994; 94US-0292968.
PF
XX
PR 22-AUG-1994; 94US-0292968.
PR 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
XX (UVAL-) UNITV ALBERTA.
XX
PI Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI Oomen RP, Read KJ, Stein PE;
DR WPI; 1999-105104/09.
XX
XX Modifications to e.g. enzymatic activity, mitogenicity and cell
PT binding of pertussis holotoxin - by identifying interaction sites of
PT a molecule with crystalline toxin and modifying the identified site
XX
XX Example 3; Fig 5; 40pp; English.
PS
XX
CC The invention relates to methods of preparing a pertussis holotoxin (PT)
CC having a modified biological activity. One method comprises identifying
CC at least 1 site in a PT that interacts with a molecule that is capable of
CC forming a complex with the holotoxin and which molecule is an effector
CC molecule which is an adenine nucleotide and which site contributes to
CC toxicity, cell binding or enzymatic activity of PT. The functional
CC interacting site(s) are identified by analysing the three dimensional
CC structure of crystalline PT, determined by X-ray crystallography. The
CC identified interacting site(s) are modified to alter toxicity, cell
CC binding or enzyme activity of the PT. The methods can be used to alter a
CC biological activity such as toxicity, enzymatic activity, mitogenicity,
CC cell binding and adjuvanticity of the PT. The three-dimensional structure
CC of PT have functional and/or structural resemblance to other bacterial
CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
CC sequence represents the beta-subunit of LT toxin.
XX
SQ Sequence 93 AA;

```

Query Match 100.0%; Score 39; DB 20; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 EYPGSOH 7
   .|||||
DB 41 EYPGSOH 47

```

Search completed: October 24, 2002, 15:35:01  
 Job time : 12.9016 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:33:23 ; Search time 4.01639 Seconds  
(without alignments)  
42.570 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPSQSH 7

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 177902

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/PCtUS.COMB.pep:\*  
7: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	93	2	US-08-292-968-26
2	39	100.0	93	2	US-08-467-974-26
3	39	100.0	93	2	US-08-467-536-26
4	39	100.0	93	3	US-08-467-976-26
5	39	100.0	93	4	US-09-082-514-26
6	31	79.5	93	4	US-09-069-023-15
7	28	71.8	39	4	US-09-450-072-78
8	28	71.8	100	4	US-09-031-962D-20
9	27	69.2	14	1	US-08-258-851-7
10	27	69.2	16	1	US-07-789-912-5
11	27	69.2	98	3	US-09-003-708A-6
12	27	69.2	98	4	US-09-383-212-6
13	26	66.7	9	4	US-09-139-802-193
14	26	66.7	11	2	US-08-747-410-12
15	26	66.7	12	1	US-08-260-582-56
16	26	66.7	12	5	PCT-US95-05471-56
17	26	66.7	14	2	US-08-467-603-58
18	26	66.7	14	2	US-08-466-793-58
19	26	66.7	14	2	US-08-491-861A-58
20	26	66.7	18	3	US-08-926-842B-56
21	26	66.7	33	1	US-08-118-906-8
22	26	66.7	33	1	US-08-486-196-8
23	26	66.7	33	1	US-08-488-135-8
24	26	66.7	33	2	US-08-474-065-8
25	26	66.7	35	1	US-08-418-893D-16
26	26	66.7	37	1	US-07-956-700B-16
27	26	66.7	37	1	US-08-476-537-16

28	26	66.7	37	1	US-08-485-607-16	Sequence 16, Appl
29	26	66.7	37	2	US-08-475-879-16	Sequence 16, Appl
30	26	66.7	95	1	US-07-956-700B-23	Sequence 23, Appl
31	26	66.7	95	1	US-08-476-537-23	Sequence 23, Appl
32	26	66.7	95	1	US-08-485-607-23	Sequence 23, Appl
33	26	66.7	95	2	US-08-475-879-23	Sequence 23, Appl
34	25	64.1	12	1	US-08-260-582-59	Sequence 59, Appl
35	25	64.1	12	5	PCT-US95-05471-59	Sequence 59, Appl
36	25	64.1	14	1	US-07-955-905A-3	Sequence 3, Appl
37	25	64.1	70	4	US-09-187-789-12	Sequence 12, Appl
38	25	64.1	73	1	US-07-956-700B-33	Sequence 33, Appl
39	25	64.1	73	1	US-08-476-537-33	Sequence 33, Appl
40	25	64.1	73	1	US-08-485-607-33	Sequence 33, Appl
41	25	64.1	73	2	US-08-475-879-33	Sequence 33, Appl
42	24.5	62.8	15	1	US-08-477-108A-11	Sequence 11, Appl
43	24	61.5	12	1	US-08-260-582-62	Sequence 62, Appl
44	24	61.5	12	5	PCT-US95-05471-62	Sequence 62, Appl
45	24	61.5	19	2	US-08-482-142-80	Sequence 80, Appl

#### ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 5856122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COHEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO.: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7  
DB 41 EVPGSOH 47

## RESULT 2

US-08-467-974-26  
; Sequence 26, Application US/08467974  
; Patent No. 5965385  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; City: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,974  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/467,536  
; FILING DATE: 06-JUN-1995  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-467-974-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7

DB 41 EVPGSOH 47

## RESULT 3

US-08-467-536-26  
; Sequence 26, Application US/08467536  
; Patent No. 5977304  
; GENERAL INFORMATION:  
; APPLICANT: READ, Randy J.  
; APPLICANT: STEIN, Penelope E.  
; APPLICANT: COCKLE, Stephen A.  
; APPLICANT: OOMEN, Raymond P.  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: KLEIN, Michel H.  
; APPLICANT: ARMSTRONG, Glen D.  
; APPLICANT: HAZES, Bart  
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; City: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,536  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,968  
; FILING DATE: 22-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/251,121  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/110,947  
; FILING DATE: 24-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I.  
; REGISTRATION NUMBER: 24,973  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 93 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-467-536-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7  
DB 41 EVPGSOH 47

## RESULT 4

US-08-467-976-26  
; Sequence 26, Application US/08467976  
; Patent No. 6018022  
; GENERAL INFORMATION:



```

: APPLICANT: READ, Randy J.
: APPLICANT: STEIN, Penelope E.
: APPLICANT: COCKLE, Stephen A.
: APPLICANT: OOMEN, Raymond P.
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: KLEIN, Michel H.
: APPLICANT: ARMSTRONG, Glen D.
: APPLICANT: HAZES, Bart
: TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467,976
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/292,968
: FILING DATE: 22-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/251,121
: FILING DATE: 31-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/110,947
: FILING DATE: 24-AUG-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-453 MIS:v9
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-467-976-26

Query Match      100.0%; Score 39; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 EVPGSOH 7
        |||||||
Db      41 EVPGSOH 47

RESULT 5
US-09-082-514-26
: Sequence 26, Application US/09082514
: Patent No. 6168928
: GENERAL INFORMATION:
: APPLICANT: READ, Randy J.
: APPLICANT: STEIN, Penelope E.
: APPLICANT: COCKLE, Stephen A.
: APPLICANT: OOMEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: APPLICANT: ARMSTRONG, Glen D.
: APPLICANT: HAZES, Bart
: TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
: NUMBER OF SEQUENCES: 46

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,514
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/292,968
: FILING DATE: 24-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-810
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-082-514-26

Query Match      100.0%; Score 39; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 EVPGSOH 7
        |||||||
Db      41 EVPGSOH 47

RESULT 6
US-09-069-023-15
: Sequence 15, Application US/09069023A
: Patent No. 6348573
: GENERAL INFORMATION:
: APPLICANT: Nunez, Gabriel
: APPLICANT: Inohara, Naohiro
: APPLICANT: Koseki, Takeyoshi
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
: FILE REFERENCE: UM-03333
: CURRENT APPLICATION NUMBER: US/09/069,023A
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 15
: LENGTH: 93
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-069-023-15

Query Match      79.5%; Score 31; DB 4; Length 93;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 VPGSOH 7
        :|||||
Db      2 MPGSOH 7

```

RESULT 7  
 US-09-450-072-78  
 ; Sequence 78, Application US/09450072  
 ; Patent No. 6358734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Delcayre, Alain  
 ; TITLE OF INVENTION: Compounds for Treatment of Infectious and Immune System Disorders  
 ; FILE REFERENCE: 11000.1042c1  
 ; CURRENT APPLICATION NUMBER: US/09/450,072  
 ; CURRENT FILING DATE: 1999-11-29  
 ; EARLIER APPLICATION NUMBER: 09/351,348  
 ; EARLIER FILING DATE: 1999-07-12  
 ; NUMBER OF SEQ ID NOS: 81  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 78  
 ; LENGTH: 39  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Made in a lab  
 US-09-450-072-78

Query Match 71.8%; Score 28; DB 4; Length 39;  
 Best Local Similarity 66.7%; Pred. No. 38;  
 Matches 4; Conservative 1; Mismatches 0; Gaps 0;

OY 2 EVPGSQ 7  
 :||| |  
 DB 30 IPGSHH 35

RESULT 8  
 US-09-031-962D-20  
 ; Sequence 20, Application US/09031962D  
 ; Patent No. 6350867  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Thomas C. Hart  
 ; APPLICANT: Jennifer A. Price  
 ; TITLE OF INVENTION: Methods and Compositions for Enhancing  
 ; TITLE OF INVENTION: Osseous Growth, Repair, and Regeneration  
 ; FILE REFERENCE: WFO98-18  
 ; CURRENT APPLICATION NUMBER: US/09/031,962D  
 ; CURRENT FILING DATE: 1998-02-27  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 20  
 ; LENGTH: 100  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-031-962D-20

Query Match 71.8%; Score 28; DB 4; Length 100;  
 Best Local Similarity 57.1%; Pred. No. 1e+02;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSQ 7  
 :|:| |  
 DB 82 EIPSEQ 88

RESULT 9  
 US-08-258-851-7  
 ; Sequence 7, Application US/08258851  
 ; Patent No. 5585244  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Elizabeth A. Allegretto  
 ; APPLICANT: J. Wesley Pike  
 ; TITLE OF INVENTION: DETECTION OF RETINOID  
 ; TITLE OF INVENTION: RECEPTOR SUBTYPE  
 ; TITLE OF INVENTION: PROTEINS  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:

ADDRESS: Lyon & Lyon  
 STREET: 611 West Sixth Street  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: USA  
 ZIP: 90017  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
 SOFTWARE: Wordperfect (Version 5.1)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/258,851  
 FILING DATE: 10 June 1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 PRIOR APPLICATION DATA: including application  
 PRIOR APPLICATION DATA: described below:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 207/199  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 14 AMINO ACIDS  
 TYPE: AMINO ACID  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: PEPTIDE  
 FEATURE:  
 OTHER INFORMATION:  
 US-08-258-851-7

Query Match 69.2%; Score 27; DB 1; Length 14;  
 Best Local Similarity 83.3%; Pred. No. 20;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQ 6  
 :||| |  
 DB 2 EVPGSQ 7

RESULT 10  
 US-07-789-912-5  
 ; Sequence 5, Application US/07789912  
 ; Patent No. 5663303  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHAMON, PIERRE  
 ; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR RAR-GAMMA OR A  
 ; TITLE OF INVENTION: SPECIFIC ISOFORM OF RAR-GAMMA  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
 STREET: 1225 Connecticut Ave. NW Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/789,912  
 FILING DATE: 19911104  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/581,694  
FILING DATE: 12-SEP-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: GOLDSTEIN, JORGE A  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 1383.0010000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..513  
US-07-789-912-5

Query Match 69.2%; Score 27; DB 1; Length 16;  
Best Local Similarity 83.3%; Pred. No. 23;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQ 6  
Db 5 EVPGSQ 10

RESULT 11  
US-09-003-708A-6  
Sequence 6, Application US/09003708A  
Patent No. 6010882  
GENERAL INFORMATION:  
APPLICANT: WU, Jen-Leih  
APPLICANT: CHEN, Jyh-Yih  
APPLICANT: CHANG, Chi-Yao  
TITLE OF INVENTION: Production of Biologically  
TITLE OF INVENTION: Active Recombinant Insulin-Like Growth Factor II  
TITLE OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Snider & Chao, L.L.P.  
STREET: P.O. Box 27613  
CITY: Washington, D.C.  
STATE: DC  
COUNTRY: USA  
ZIP: 20038  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/003,708A  
FILING DATE: 01/07/98  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/034/736  
FILING DATE: January 10, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ronald R. Snider  
REGISTRATION NUMBER: 24,962  
REFERENCE/DOCKET NUMBER: ACA-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)-347-2600  
TELEFAX: (202)-347-7870  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: Unknown

MOLECULE TYPE: polypeptide  
HYPOTHETICAL: NO  
US-09-003-708A-6

Query Match 69.2%; Score 27; DB 3; Length 98;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSQ 7  
Db 20 EVPRKQH 26

RESULT 12  
US-09-383-212-6  
Sequence 6, Application US/09383212  
Patent No. 6235874  
GENERAL INFORMATION:  
APPLICANT: WU, Jen-Leih  
APPLICANT: CHEN, Jyh-Yih  
TITLE OF INVENTION: Production of Biologically Active  
TITLE OF INVENTION: Recombinant Insulin-Like Growth Factor II  
TITLE OF INVENTION: Polypeptides  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Snider & Chao, L.L.P.  
STREET: P.O. Box 27613  
CITY: Washington, D.C.  
STATE: DC  
COUNTRY: USA  
ZIP: 20038  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/383,212  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/003,708  
FILING DATE: 01/07/98  
APPLICATION NUMBER: 60/034/736  
FILING DATE: January 10, 1997  
APPLICATION NUMBER: 09/003/708  
FILING DATE: January 7, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Fel-Fel Chao  
REGISTRATION NUMBER: 43,538  
REFERENCE/DOCKET NUMBER: ACA-001-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)-347-2600  
TELEFAX: (202)-347-7870  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: Unknown  
MOLECULE TYPE: polypeptide  
HYPOTHETICAL: NO  
US-09-383-212-6

Query Match 69.2%; Score 27; DB 4; Length 98;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSQ 7  
Db 20 EVPRKQH 26

RESULT 13  
US-09-139-802-193  
Sequence 193, Application US/09139802  
Patent No. 6180084  
GENERAL INFORMATION:  
APPLICANT: Ruoslahti, Erkki  
TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
TITLE OF INVENTION: Same  
FILE REFERENCE: P-LJ 3203  
CURRENT APPLICATION NUMBER: US/09/139,802  
EARLIER FILING DATE: 1998-08-25  
EARLIER FILING DATE: 1997-09-10  
EARLIER APPLICATION NUMBER: 08/710,067  
EARLIER FILING DATE: 1996-09-10  
NUMBER OF SEQ ID NOS: 226  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 193  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-139-802-193  
Query Match  
Best Local Similarity 83.3%; Score 26; DB 4; Length 9;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 EYPGSC 6  
Db 4 EYPGSC 9  
RESULT 14  
US-08-747-410-12  
Sequence 12, Application US/08747410  
Patent No. 5993820  
GENERAL INFORMATION:  
APPLICANT: BAGDASARIAN, Michael  
TITLE OF INVENTION: CHIMERIC LTB VACCINES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
STREET: 3100 No. 5993820west Center, 90 South Seventh St  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/747,410  
FILING DATE: 12-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ketelberger, Denise M  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 11526.1-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612/371-5268  
TELEFAX: 612/332-9081  
TELEX:

INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
US-08-747-410-12  
Query Match  
Best Local Similarity 66.7%; Score 26; DB 2; Length 11;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EYPGS 5  
Db 2 EYPGS 6  
RESULT 15  
US-08-260-582-56  
Sequence 56, Application US/08260582  
Patent No. 5635182  
GENERAL INFORMATION:  
APPLICANT: McCoy, John M.  
TITLE OF INVENTION: METHOD OF DETECTING LIGAND  
TITLE OF INVENTION: INTERACTIONS  
NUMBER OF SEQUENCES: 76  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/260,582  
FILING DATE: 16-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meinerly, M. C.  
REGISTRATION NUMBER: 31,544  
REFERENCE/DOCKET NUMBER: GI 5236  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-260-582-56  
Query Match  
Best Local Similarity 66.7%; Score 26; DB 1; Length 12;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VPGSQ 6  
Db 3 VPGSQ 7

Fri Oct 25 17:50:44 2002

us-09-786-648-2.lenlim.ra

Page 7

Search completed: October 24, 2002, 15:38:21  
Job time : 5.01639 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:09 ; Search time 46.5902 Seconds  
(without alignments)  
52.884 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 2671256

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_MA Main:\*

1: /cgn2\_6/ptodata/1/paa/PCNUS.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US081.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US080.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082.COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US60.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	100.0	7	21	US-09-786-648-2	Sequence 2, Appli
2	39	100.0	12	21	US-09-786-648-3	Sequence 3, Appli
3	39	100.0	15	11	US-08-732-371-1	Sequence 1, Appli
4	39	100.0	15	11	US-08-732-371A-1	Sequence 1, Appli
5	39	100.0	21	21	US-09-786-648-4	Sequence 4, Appli
6	39	100.0	21	21	US-09-786-648-5	Sequence 5, Appli
7	39	100.0	93	5	US-08-110-947-10	Sequence 10, Appli

8	39	100.0	93	5	US-08-110-947A-26	Sequence 26, Appli
9	39	100.0	93	6	US-08-251-121-26	Sequence 26, Appli
10	34	87.2	41	20	US-09-688-051-2967	Sequence 2967, Ap
11	34	87.2	84	20	US-09-617-682A-5651	Sequence 5651, Ap
12	33	84.6	51	22	US-09-866-066-35	Sequence 35, Appli
13	32	82.1	64	26	US-60-192-587-1036	Sequence 1036, Ap
14	32	82.1	64	26	US-60-194-091-1018	Sequence 1018, Ap
15	31	79.5	41	12	US-08-888-755-16	Sequence 16, Appli
16	31	79.5	41	12	US-08-888-755A-16	Sequence 16, Appli
17	31	79.5	61	24	US-10-029-386-34288	Sequence 34288, A
18	31	79.5	64	21	US-09-758-446-1554	Sequence 1554, Ap
19	31	79.5	75	1	PCT-US01-14827-12024	Sequence 12024, A
20	31	79.5	75	18	US-09-465-854-18	Sequence 18, Appli
21	31	79.5	75	26	US-60-112-736-8	Sequence 8, Appli
22	31	79.5	91	19	US-09-595-238A-470	Sequence 470, App
23	30	76.9	52	26	US-60-178-308-2956	Sequence 2956, Ap
24	30	76.9	69	21	US-09-757-034-752	Sequence 752, App
25	30	76.9	71	26	US-60-173-465-440	Sequence 440, App
26	30	76.9	77	20	US-09-688-051-1575	Sequence 1575, Ap
27	30	76.9	79	13	US-08-905-279-481	Sequence 481, App
28	30	76.9	82	26	US-60-186-656-1008	Sequence 1008, Ap
29	30	76.9	82	26	US-60-192-739-2951	Sequence 2951, Ap
30	30	76.9	82	26	US-60-194-243-2113	Sequence 2113, Ap
31	30	76.9	84	1	PCT-US01-01354-12253	Sequence 12253, A
32	30	76.9	84	21	US-09-764-905-12253	Sequence 12253, A
33	30	76.9	84	24	US-10-092-399-12253	Sequence 12253, A
34	30	76.9	95	21	US-09-764-874-6620	Sequence 6620, App
35	30	76.9	90	26	US-60-188-162-4087	Sequence 4087, App
36	30	76.9	90	26	US-60-173-465-530	Sequence 530, App
37	30	76.9	90	26	US-60-192-739-2952	Sequence 2952, App
38	30	76.9	93	13	US-08-905-144-352	Sequence 352, App
39	30	76.9	95	1	PCT-US01-01334-6620	Sequence 6620, App
40	30	76.9	95	21	US-09-764-874-6620	Sequence 6620, App
41	30	76.9	96	21	US-09-760-495-583	Sequence 583, App
42	30	76.9	97	26	US-60-200-109-1132	Sequence 1132, App
43	30	76.9	98	26	US-60-200-109-1129	Sequence 1129, App
44	29	74.4	26	1	PCT-US98-23435-394	Sequence 394, App
45	29	74.4	26	17	US-09-305-736-488	Sequence 488, App

#### ALIGNMENTS

RESULT 1  
US-09-786-648-2  
; Sequence 2, Application US/09786648  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Neil Andrew  
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vacci  
; FILE REFERENCE: 7438  
; CURRENT APPLICATION NUMBER: US/09/786,648  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: PCT/GB99/02970  
; PRIORITY FILING DATE: 1999-09-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: MS DOS  
; SEQ ID NO 2  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: E. coli  
; FEATURE:  
; LOCATION: 51...57  
; OTHER INFORMATION: Isolated or synthetic Etxb beta4-alpha2 loop fragment derivabl  
; OTHER INFORMATION: human variant E. coli  
US-09-786-648-2

Query Match  
Best Local Similarity 100.0%; Score 39; DB 21; Length 7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EVPGSQH 7

Db 1 EVPGSOH 7

|||||

RESULT 2

US-09-786-648-3

Sequence 3, Application US/09786648

GENERAL INFORMATION:

APPLICANT: Williams, Neil Andrew

APPLICANT: Hirst, Timothy Raymond

TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine

FILE REFERENCE: 7438

CURRENT APPLICATION NUMBER: US/09/786,648

CURRENT FILING DATE: 2001-03-07

PRIOR APPLICATION NUMBER: PCT/GB99/02970

PRIOR FILING DATE: 1999-09-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: MS DOS

SEQ ID NO: 3

LENGTH: 12

TYPE: PRT

ORGANISM: E. coli

FEATURE:

LOCATION: 50..61

OTHER INFORMATION: Isolated or synthetic EtXB beta4-alpha2 loop fragment derivable

US-09-786-648-3

Query Match

Best Local Similarity 100.0%; Score 39; DB 21; Length 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7

|||||

Db 2 EVPGSOH 8

RESULT 3

US-08-732-371-1

Sequence 1, Application US/08732371

GENERAL INFORMATION:

APPLICANT: MIRELMAN, David

APPLICANT: MARKS, Robert S.

APPLICANT: SELA, Michael

TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST

TITLE OF INVENTION: INFECTING AGENTS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/732,371

FILING DATE: 09-JAN-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 109519

FILING DATE: 03-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: MIRELMAN-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-732-371-1

Query Match

Best Local Similarity 100.0%; Score 39; DB 11; Length 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7

|||||

Db 2 EVPGSOH 8

RESULT 4

US-08-732-371A-1

Sequence 1, Application US/08732371A

GENERAL INFORMATION:

APPLICANT: MIRELMAN, David

APPLICANT: MARKS, Robert S.

APPLICANT: SELA, Michael

TITLE OF INVENTION: VACCINES FOR ORAL IMMUNIZATION AGAINST

TITLE OF INVENTION: INFECTING AGENTS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/732,371A

FILING DATE: 09-JAN-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IL 109519

FILING DATE: 03-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971

REFERENCE/DOCKET NUMBER: MIRELMAN-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-732-371A-1

Query Match

Best Local Similarity 100.0%; Score 39; DB 11; Length 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7

|||||

Db 2 EVPGSOH 8

RESULT 5

US-09-786-648-4



```
; Sequence 4, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 4
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic Ecxb beta4-alpha2 loop fragment derivable f
US-09-786-648-4
;
Query Match          100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EVPGSOH 7
       1111111
Db      7 EVPGSOH 13

RESULT 5
US-09-786-648-5
; Sequence 5, Application US/09786648
; GENERAL INFORMATION:
; APPLICANT: Williams, Neil Andrew
; TITLE OF INVENTION: Peptide Fragments of Cholera Toxin B or Enterotoxin B as Vaccine
; FILE REFERENCE: 7438
; CURRENT APPLICATION NUMBER: US/09/786,648
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: PCT/GB99/02970
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: MS DOS
; SEQ ID NO 5
; LENGTH: 21
; TYPE: PRT
; ORGANISM: E. coli
; FEATURE:
; LOCATION: 45...65
; OTHER INFORMATION: isolated or synthetic Ecxb beta4-alpha2 loop fragment derivable f
US-09-786-648-5
;
Query Match          100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EVPGSOH 7
       1111111
Db      7 EVPGSOH 13

RESULT 7
US-08-110-947-10
; Sequence 10, Application US/08110947
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
```

```
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947
; FILING DATE: 24-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FALLOW, Charles W
; REGISTRATION NUMBER: 28,946
; REFERENCE/DOCKET NUMBER: 1038-303 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: 89-9456 LUKPAT
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-110-947-10
;
Query Match          100.0%; Score 39; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EVPGSOH 7
       1111111
Db      41 EVPGSOH 47

RESULT 8
US-08-110-947A-26
; Sequence 26, Application US/08110947A
; GENERAL INFORMATION:
; APPLICANT: READ, Randy J
; APPLICANT: STEIN, Penelope E
; APPLICANT: COCKLE, Stephen A
; APPLICANT: COHEN, Raymond P
; APPLICANT: KLEIN, Michel H
; TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: Suite 1203 Crystal Plaza I, 2001 Jefferson
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,947A
; FILING DATE: 24-AUG-1993
```

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLON, Charles W  
REGISTRATION NUMBER: 28, 946  
REFERENCE/DOCKET NUMBER: 1038-303 MIS-Jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0810  
TELEX: 89-9456 LUKPAT  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-110-947A-26

Query Match 100.0%; Score 39; DB 5; Length 93;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
DB 41 EVPGSQH 47

RESULT 9  
US-08-251-121-26  
Sequence 26, Application US/08251121  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/251.121  
FILING DATE: 31-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-335  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-251-121-26

Query Match 100.0%; Score 39; DB 6; Length 93;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
DB 41 EVPGSQH 47

RESULT 10  
US-09-688-051-2967  
Sequence 2967, Application US/09688051  
GENERAL INFORMATION:  
APPLICANT: ALEXANDROV, Nikolai  
APPLICANT: BROVER, Vyacheslav  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid  
TITLE OF INVENTION: Thereby  
FILE REFERENCE: 2750-1242P  
CURRENT APPLICATION NUMBER: US/09/688,051  
CURRENT FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 4193  
SEQ ID NO 2967  
LENGTH: 41  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..41  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1..41  
OTHER INFORMATION: Ceres Seq. ID 1714257  
US-09-688-051-2967

Query Match 87.2%; Score 34; DB 20; Length 41;  
Best Local Similarity 71.4%; Pred. No. 78;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
DB 9 KIPGSQH 15

RESULT 11  
US-09-617-682A-5651  
Sequence 5651, Application US/09617682A  
GENERAL INFORMATION:  
APPLICANT: ALEXANDROV, Nikolai  
APPLICANT: BROVER, Vyacheslav  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid  
TITLE OF INVENTION: Thereby  
FILE REFERENCE: 2750-1063P  
CURRENT APPLICATION NUMBER: US/09/617,682A  
CURRENT FILING DATE: 2000-07-19  
NUMBER OF SEQ ID NOS: 16871  
SEQ ID NO 5651  
LENGTH: 84  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..84  
OTHER INFORMATION: Xaa is any amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1..84  
OTHER INFORMATION: Ceres Seq. ID 1355962  
US-09-617-682A-5651

Query Match 87.2%; Score 34; DB 20; Length 84;  
Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7

Db 52 KIPGSH 58

:::|||||

RESULT 12  
US-09-866-066-35  
; Sequence 35, Application US/09866066  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Christopher  
; APPLICANT: Roberds, Steve  
; APPLICANT: Ruble, Cara  
; APPLICANT: Gotow, Lisa  
; APPLICANT: Karnovsky, Alla  
; TITLE OF INVENTION: Human Ion Channels  
; FILE REFERENCE: 00229.US1  
; CURRENT APPLICATION NUMBER: US/09/866,066  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/207,152  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/207,257  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/207,119  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 35  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-866-066-35

Query Match 84.6%; Score 33; DB 22; Length 51;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSH 7  
|:|:|:|  
Db 8 EIPGSH 14

RESULT 13  
US-60-192-587-1036  
; Sequence 1036, Application US/60192587  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien  
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; FILE REFERENCE: C1000405  
; CURRENT APPLICATION NUMBER: US/60/192,587  
; CURRENT FILING DATE: 2000-03-28  
; NUMBER OF SEQ ID NOS: 1666  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1036  
; LENGTH: 64  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-60-192-587-1036

Query Match 82.1%; Score 32; DB 26; Length 64;  
Best Local Similarity 71.4%; Pred. No. 3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSH 7  
|:|:|:|  
Db 42 EIPGSH 48

RESULT 14  
US-60-194-091-1018  
; Sequence 1018, Application US/60194091  
; GENERAL INFORMATION:  
; APPLICANT: Bonazzi, Vivien

; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; FILE REFERENCE: C1000416  
; CURRENT APPLICATION NUMBER: US/60/194,091  
; CURRENT FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 1598  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1018  
; LENGTH: 64  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-60-194-091-1018

Query Match 82.1%; Score 32; DB 26; Length 64;  
Best Local Similarity 71.4%; Pred. No. 3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSH 7  
|:|:|:|  
Db 42 EIPGSH 48

RESULT 15  
US-08-888-755-16  
; Sequence 16, Application US/0888755  
; GENERAL INFORMATION:  
; APPLICANT: Russell, John  
; APPLICANT: Colpitts, Tracey  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; TITLE OF INVENTION: FOR DETECTING DISEASE OF THE BREAST  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/888,755  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6129.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 41 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: None  
US-08-888-755-16

Query Match 79.5%; Score 31; DB 12; Length 41;  
Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSH 7

Db :|||||  
32 MPGSQH 37

Search completed: October 24, 2002, 15:45:14  
Job time : 48.5902 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:35:39 : Search time 15.2623 Seconds  
(without alignments)  
137.008 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EPPGSH 7

Scoring table: BLOSUM62

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 378271

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PC/PC\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	79.5	64	US-10-212-759-1554	Sequence 1554, App
2	31	79.5	82	PCT-US02-09921-833	Sequence 833, App
3	30	76.9	69	US-10-235-954-752	Sequence 752, App
4	30	76.9	79	US-09-547-559C-2880	Sequence 2880, App
5	30	76.9	93	US-09-547-559C-2428	Sequence 2428, App
6	30	76.9	96	US-10-143-775-583	Sequence 583, App
7	29	74.4	29	US-10-182-995-21108	Sequence 21108, App
8	29	74.4	29	US-10-203-134-27374	Sequence 27374, App
9	29	74.4	29	US-10-203-136-27380	Sequence 27380, App
10	29	74.4	29	US-10-182-993-26770	Sequence 26770, App
11	29	74.4	29	US-10-182-997-19756	Sequence 19756, App
12	29	74.4	29	US-10-182-998-11398	Sequence 11398, App
13	29	74.4	29	US-10-203-135-26391	Sequence 26391, App
14	29	74.4	29	US-10-203-137-27628	Sequence 27628, App
15	29	74.4	29	US-10-203-138-11687	Sequence 11687, App
16	29	74.4	29	US-10-203-139-26533	Sequence 26533, App
17	29	74.4	86	US-10-195-730-115	Sequence 115, App
18	29	74.4	86	US-10-195-730-216	Sequence 216, App
19	28	71.8	33	US-10-106-698-7689	Sequence 7689, App
20	28	71.8	39	US-10-100-679-78	Sequence 78, App
21	28	71.8	55	US-09-620-393B-4797	Sequence 4797, App
22	28	71.8	58	US-09-807-512-10	Sequence 10, App
23	28	71.8	70	US-10-222-911-2237	Sequence 2237, App
24	28	71.8	73	US-10-235-953-3603	Sequence 3603, App
25	28	71.8	78	US-10-182-995-28650	Sequence 28650, App
26	28	71.8	80	US-10-219-999-33961	Sequence 33961, App

27	28	71.8	87	US-10-155-881-12211	Sequence 12211, App
28	28	71.8	97	US-10-235-953-3641	Sequence 3641, App
29	27	69.2	19	US-10-062-831-123	Sequence 123, App
30	27	69.2	29	US-10-106-698-8197	Sequence 8197, App
31	27	69.2	30	US-10-106-698-8604	Sequence 8604, App
32	27	69.2	31	US-10-106-698-6625	Sequence 6607, App
33	27	69.2	32	US-10-106-698-6607	Sequence 6612, App
34	27	69.2	33	US-10-106-698-6637	Sequence 6637, App
35	27	69.2	33	US-10-106-698-7879	Sequence 7879, App
36	27	69.2	33	US-10-040-394-15	Sequence 15, App
37	27	69.2	34	US-10-106-698-5746	Sequence 5746, App
38	27	69.2	34	US-10-106-698-6618	Sequence 6618, App
39	27	69.2	34	US-10-106-698-8196	Sequence 8196, App
40	27	69.2	34	US-10-106-698-8458	Sequence 8458, App
41	27	69.2	35	US-10-106-698-8634	Sequence 8634, App
42	27	69.2	35	US-10-106-698-8340	Sequence 8340, App
43	27	69.2	36	US-10-106-698-6626	Sequence 6626, App
44	27	69.2	36	US-10-106-698-6636	Sequence 6636, App
45	27	69.2	36	US-10-106-698-6636	Sequence 6636, App

## ALIGNMENTS

RESULT 1  
US-10-212-759-1554  
Sequence 1554, Application US/10212759  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM019CIN  
CURRENT APPLICATION NUMBER: US/10/212,759  
CURRENT FILING DATE: 2002-08-07  
PRIOR APPLICATION NUMBER: 09/758,446  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
NUMBER OF SEQ ID NOS: 1734  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 1554  
LENGTH: 64  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (16)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (17)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-212-759-1554  
Query Match  
Best Local Similarity 79.5%; Score 31; DB 6; Length 64;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 VPGSH 7  
|||||  
DB 54 VPGSH 59  
RESULT 2  
PCT-US02-09921-833  
Sequence 833, Application PC/TUS0209921  
GENERAL INFORMATION:  
APPLICANT: INCYTE GENOMICS, INC.

```

; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Ping B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEEFELD, Yael
; APPLICANT: GERSTIN JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 83
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: LI:036747.17.orf2:2001MAY17
; PCT-US02-09921-833

Query Match          79.5%; Score 31; DB 1; Length 82;
Best Local Similarity 83.3%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 VPGSOH 7
      111111
Db      24 VPGTOH 29

RESULT 3
US-10-235-954-752
; Sequence 752, Application US/10235954
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM002CIN
; CURRENT APPLICATION NUMBER: US/10/235,954
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/757,034
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1440

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 752
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-235-954-752

Query Match          76.9%; Score 30; DB 6; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 PGSOH 7
      111111
Db      57 PGSOH 61

RESULT 4
US-09-547-599C-2880
; Sequence 2880, Application US/09547599C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; FILE REFERENCE: GEN-T119C1
; CURRENT APPLICATION NUMBER: US/09/547,599C
; CURRENT FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 08/905,223
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,135
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,051
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,144
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: Patent.pm
; SEQ ID NO 2880
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Ovary
; NAME/KEY: SIGNAL
; LOCATION: -55..-1
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 4.60
; OTHER INFORMATION: seq CALSLDPARGAGC/GR
; US-09-547-599C-2880

Query Match          76.9%; Score 30; DB 5; Length 79;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 VPGSOH 7
      111111
Db      63 VPGSOH 68

RESULT 5
US-09-547-599C-2428
; Sequence 2428, Application US/09547599C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste

```

```

: APPLICANT: Duclert, Aymeric
: TITLE OF INVENTION: Lactox, Bruno
: FILE REFERENCE: GEN-T119C1
: CURRENT APPLICATION NUMBER: US/09/547,599C
: PRIOR FILING DATE: 2000-04-11
: PRIOR APPLICATION NUMBER: US 08/905,223
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,135
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,051
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,144
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,279
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/904,468
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,134
: PRIOR FILING DATE: 1997-08-01
: PRIOR APPLICATION NUMBER: US 08/905,133
: PRIOR FILING DATE: 1997-08-01
: NUMBER OF SEQ ID NOS: 3475
: SOFTWARE: Patent.pm
: SEQ ID NO 2428
: LENGTH: 93
: TYPE: PRT
: ORGANISM: Homo Sapiens
: FEATURE:
: OTHER INFORMATION: Normal prostate
: NAME/KEY: SIGNAL
: LOCATION: -60..-1
: OTHER INFORMATION: Von Heljne matrix
: OTHER INFORMATION: score 7.00
: FEATURE:
: OTHER INFORMATION: seq VLVVLAIRSLGRS/CS
: NAME/KEY: UNSURE
: LOCATION: 30
: OTHER INFORMATION: Xaa - *,Tyr
: US-09-547-599C-2428

Query Match          76.9%; Score 30; DB 5; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 PGSQH 7
      11111
Db      81 PGSQH 85

RESULT 6
US-10-143-775-583
: Sequence 583, Application US/10143775
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC021C1N
: CURRENT APPLICATION NUMBER: US/10/143,775
: CURRENT FILING DATE: 2002-05-14
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 1064
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 583
: LENGTH: 96
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-143-775-583

Query Match          76.9%; Score 30; DB 6; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      3 PGSQH 7
      11111
Db      76 PGSQH 80

RESULT 7
US-10-182-995-21108
: Sequence 21108, Application US/10182995
: GENERAL INFORMATION:
: APPLICANT: Molecular Dynamics, Inc.
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: PB 0004 WO 1
: CURRENT APPLICATION NUMBER: US/10/182,995
: CURRENT FILING DATE: 2002-08-02
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 04 February 2000 (04.02.00)
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 26 May 2000 (26.05.00)
: PRIOR APPLICATION NUMBER: US 09/632,365
: PRIOR FILING DATE: 03 August 2000 (03.08.00)
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 03 October 2000 (03.10.00)
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 27 September 2000 (27.09.00)
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 21 September 2000 (21.09.00)
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 30 June 2000 (30.06.00)
: NUMBER OF SEQ ID NOS: 29119
: SOFTWARE: Molecular Dynamics Sequence Listing Engine
: SEQ ID NO 21108
: LENGTH: 29
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC009948.2
: FEATURE:
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
: FEATURE:
: OTHER INFORMATION: EST_HUMAN HIT: BF063046.1, EVALUOE 1.00e-09
: FEATURE:
: OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUOE 1.60e+00
: US-10-182-995-21108

Query Match          74.4%; Score 29; DB 6; Length 29;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 VPGSH 7
      11111
Db      4 IPGSH 9

RESULT 8
US-10-203-134-27374
: Sequence 27374, Application US/10203134
: GENERAL INFORMATION:
: APPLICANT: Molecular Dynamics, Inc.
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: PB 0004 WO 6
: CURRENT APPLICATION NUMBER: US/10/203,134
: CURRENT FILING DATE: 2002-08-02
```

PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 04 February 2000 (04.02.00)  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 26 May 2000 (26.05.00)  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 03 August 2000 (03.08.00)  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 03 October 2000 (03.10.00)  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 27 September 2000 (27.09.00)  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 21 September 2000 (21.09.00)  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 30 June 2000 (30.06.00)  
NUMBER OF SEQ ID NOS: 38628  
SOFTWARE: Molecular Dynamics Sequence Listing Engine  
SEQ ID NO 27374  
LENGTH: 29  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009948.2  
FEATURE:  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
FEATURE:  
OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
FEATURE:  
OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-203-134-27374

Query Match  
Best Local Similarity 74.4%; Score 29; DB 6; Length 29;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 4 IPGEQH 9

QY 2 VPGEQH 7  
: || ||  
DB 4 IPGEQH 9

RESULT 9  
US-10-203-136-27380  
GENERAL INFORMATION:  
APPLICANT: Molecular Dynamics, Inc.  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: PB 0004 WO 3  
CURRENT APPLICATION NUMBER: US/10/203,136  
CURRENT FILING DATE: 2002-08-02  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 04 February 2000 (04.02.00)  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 26 May 2000 (26.05.00)  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 03 August 2000 (03.08.00)  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 03 October 2000 (03.10.00)  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 27 September 2000 (27.09.00)  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 21 September 2000 (21.09.00)  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 30 June 2000 (30.06.00)  
NUMBER OF SEQ ID NOS: 38578  
SOFTWARE: Molecular Dynamics Sequence Listing Engine  
SEQ ID NO 27380  
LENGTH: 29  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
OTHER INFORMATION: MAP TO AC009948.2  
FEATURE:  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
FEATURE:  
OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
FEATURE:  
OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-203-136-27380

Query Match  
Best Local Similarity 74.4%; Score 29; DB 6; Length 29;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 4 IPGEQH 9

QY 2 VPGEQH 7  
: || ||  
DB 4 IPGEQH 9

RESULT 10  
US-10-182-993-26770  
GENERAL INFORMATION:  
APPLICANT: Molecular Dynamics, Inc.  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: PB 0004 WO 2  
CURRENT APPLICATION NUMBER: US/10/182,993  
CURRENT FILING DATE: 2002-08-02  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 04 February 2000 (04.02.00)  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 26 May 2000 (26.05.00)  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 03 August 2000 (03.08.00)  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 03 October 2000 (03.10.00)  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 27 September 2000 (27.09.00)  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 21 September 2000 (21.09.00)  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 30 June 2000 (30.06.00)  
NUMBER OF SEQ ID NOS: 37811  
SOFTWARE: Molecular Dynamics Sequence Listing Engine  
SEQ ID NO 26770  
LENGTH: 29  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009948.2  
FEATURE:  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
FEATURE:  
OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
FEATURE:  
OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-182-993-26770

Query Match  
Best Local Similarity 74.4%; Score 29; DB 6; Length 29;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
DB 4 IPGEQH 9

QY 2 VPGEQH 7  
: || ||  
DB 4 IPGEQH 9

RESULT 11



US-10-182-997-19756  
; Sequence 19756, Application US/10182997  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HELA  
; FILE REFERENCE: PB 0004 WO 10  
; CURRENT APPLICATION NUMBER: US/10/182,997  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
; NUMBER OF SEQ ID NOS: 26941  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 19756  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: MAP TO AC009948.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
; FEATURE:  
; OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
; FEATURE:  
; OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-182-997-19756  
Query Match 74.4%; Score 29; DB 6; Length 29;  
Best Local Similarity 66.7%; Pred. No. 79;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VPGSOH 7  
; 11 11  
DB 4 IPGEQH 9  
RESULT 12  
US-10-182-998-11398  
; Sequence 11398, Application US/10182998  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HBL 100  
; FILE REFERENCE: PB 0004 WO 9  
; CURRENT APPLICATION NUMBER: US/10/182,998  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
; NUMBER OF SEQ ID NOS: 15009  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 11398  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: MAP TO AC009948.2  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6  
; FEATURE:  
; OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
; FEATURE:  
; OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-182-998-11398  
Query Match 74.4%; Score 29; DB 6; Length 29;  
Best Local Similarity 66.7%; Pred. No. 79;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VPGSOH 7  
; 11 11  
DB 4 IPGEQH 9  
RESULT 13  
US-10-203-135-26391  
; Sequence 26391, Application US/10203135  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN LUNG  
; FILE REFERENCE: PB 0004 WO 5  
; CURRENT APPLICATION NUMBER: US/10/203,135  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 30 June 2000 (30.06.00)  
; NUMBER OF SEQ ID NOS: 37012  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 26391  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE: MAP TO AC009948.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
; FEATURE:  
; OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
; FEATURE:

OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-203-138-26391

Query Match 74.4%; Score 29; DB 6; Length 29;  
Best Local Similarity 66.7%; Pred. No. 79;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPSSQH 7  
:|||  
DB 4 IPEQH 9

RESULT 14  
US-10-203-137-27628  
; Sequence 27628, Application US/10203137  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: PB 0004 WO 7  
; CURRENT APPLICATION NUMBER: US/10/203,137  
; PRIOR FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; NUMBER OF SEQ ID NOS: 38837  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 27628

LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC009948.2  
; FEATURE:  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
; FEATURE:  
; OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
; OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-203-137-27628

Query Match 74.4%; Score 29; DB 6; Length 29;  
Best Local Similarity 66.7%; Pred. No. 79;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPSSQH 7  
:|||  
DB 4 IPEQH 9

RESULT 15  
US-10-203-138-11687  
; Sequence 11687, Application US/10203138  
; GENERAL INFORMATION:  
; APPLICANT: Molecular Dynamics, Inc.  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: PB 0004 WO 8  
; CURRENT APPLICATION NUMBER: US/10/203,138  
; CURRENT FILING DATE: 2002-08-02  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 04 February 2000 (04.02.00)  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 26 May 2000 (26.05.00)  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 03 August 2000 (03.08.00)  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 03 October 2000 (03.10.00)  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 27 September 2000 (27.09.00)  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 21 September 2000 (21.09.00)  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; NUMBER OF SEQ ID NOS: 15438  
; SOFTWARE: Molecular Dynamics Sequence Listing Engine  
; SEQ ID NO 11687

LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC009948.2  
; FEATURE:  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4  
; OTHER INFORMATION: EST\_HUMAN HIT: BF063046.1, EVALUE 1.00e-09  
; OTHER INFORMATION: SWISSPROT HIT: P16258, EVALUE 1.60e+00  
US-10-203-138-11687

Query Match 74.4%; Score 29; DB 6; Length 29;  
Best Local Similarity 66.7%; Pred. No. 79;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPSSQH 7  
:|||  
DB 4 IPEQH 9

Search completed: October 24, 2002, 15:47:33  
Job time : 16.2623 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:31:03 : Search time 5.27869 seconds  
(without alignments)  
127.423 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 37652

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR\_71:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	71.8	97	2	AH3516
2	27	69.2	75	2	S25719
3	27	69.2	87	2	T06281
4	27	69.2	87	2	T01665
5	27	69.2	93	2	T20227
6	26	66.7	66	2	D83589
7	26	66.7	73	4	Q0ECW2
8	26	66.7	74	2	G83393
9	26	66.7	84	2	T17876
10	26	66.7	84	2	D83940
11	26	66.7	96	2	AF3175
12	26	66.7	97	2	E70728
13	26	66.7	99	2	JV0104
14	25	64.1	20	2	A39328
15	25	64.1	33	2	A60222
16	25	64.1	52	4	S70252
17	25	64.1	72	2	D98143
18	25	64.1	76	2	T26894
19	25	64.1	81	1	IHM7
20	25	64.1	97	2	AH2175
21	24	61.5	34	2	S78590
22	24	61.5	39	2	T45362
23	24	61.5	61	2	A39327
24	24	61.5	67	2	T07555
25	24	61.5	72	2	C36760
26	24	61.5	73	2	A49550
27	24	61.5	74	2	C49530
28	24	61.5	78	2	G97208
29	24	61.5	79	2	S77831

30	24	61.5	87	2	T00775	hypothetical prote
31	24	61.5	88	2	S61552	mammery tumor prot
32	24	61.5	96	2	T45337	hypothetical prote
33	24	61.5	97	2	H42409	glutathione transf
34	24	61.5	100	2	C72728	hypothetical prote
35	23.5	60.3	92	2	I50644	gonadolibertin I pr
36	23	59.0	32	2	A53505	pyruvate,orthophos
37	23	59.0	34	2	C31514	hemopexin - chicke
38	23	59.0	35	2	G60529	hemocyanin M3' - c
39	23	59.0	40	2	F45095	photosystem I ligh
40	23	59.0	44	2	S40297	phosphoenolpyruvat
41	23	59.0	47	2	S69145	gamma-thionin SI-a
42	23	59.0	47	2	B96949	hypothetical prote
43	23	59.0	55	1	R5H832	ribosomal protein
44	23	59.0	55	1	H87442	hypothetical prote
45	23	59.0	58	2	AC1827	hypothetical prote

#### ALIGNMENTS

RESULT 1  
AH3516  
hypothetical protein BMEI10058 [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AH3516  
R:DeVecchio, V.G.; Kapatal, V.; Redhar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov, J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AH3516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <KUR>  
A:Cross-references: GB:AE008918; PIDN:ALU53299.1; PID:g17984183; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10058  
A:Map position: II

Query Match  
Best Local Similarity 71.8%; Score 28; DB 2; Length 97;  
Matches 4; Conservative 2; Mismatches 0; Gaps 0;  
Indels 0;  
Db 32 VPGSQH 37

RESULT 2  
S25719  
protein ECL - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 05-Nov-1999  
C:Accession: S25719  
R:Shen, P.; Howlett, G.J.  
Arch. Biochem. Biophys. 297, 345-353, 1992  
A:Title: Two coding regions closely linked to the rat apolipoprotein E gene: nucleot  
A:Reference number: S25719; MUID:92359549  
A:Accession: S25719  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-75 <SHD>  
A:Cross-references: EMBL:X56190; NID:g56075; PIDN:CAA39653.1; PID:g56076

Query Match  
Best Local Similarity 69.2%; Score 27; DB 2; Length 75;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 2 VPGSQH 7

DB 25 VPGSGH 30

# RESULT 3

reverse transcriptase - tomato retrotransposon Tyl-copia class (fragment)  
 C:Species: lycopersicon esculentum (tomato)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 08-Oct-1999  
 C:Accession: T06281  
 R:Rogers, S.A.; Pauls, K.P.  
 A:Description: Tyl-copia class retrotransposons of tomato.  
 A:Reference number: Z15586  
 A:Accession: T06281  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-87 <ROG>  
 A:Cross-references: EMBL:AF072638; NID:g3550085; PIDN:ANC34605.1; PID:g3550086  
 A:Experimental source: cultivar H72; tissue-type leaf  
 C:Superfamily: retrovirus-related polyprotein

Query Match 69.2%; Score 27; DB 2; Length 87;  
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPGSGH 7  
 DB 22 VPGKEH 27

# RESULT 4

T01665  
 tat protein - human immunodeficiency virus type 1  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T01665  
 R:Allison, M.; Main-Hobson, S.; Gluckman, J.C.; Sonigo, P.  
 Cell 46, 63-74, 1986  
 A:Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol  
 A:Reference number: Z14389; MUID:86245056  
 A:Accession: T01665  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-87 <ALT>  
 A:Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28015.1; PID:g60233  
 C:Genetics:  
 A:Introns: 72/2  
 C:Superfamily: AIDS trans-activating transcription regulator

Query Match 69.2%; Score 27; DB 2; Length 87;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EYVGSOH 7  
 DB 71 EOPSSOH 77

# RESULT 5

T20227  
 hypothetical protein C54G10.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T20227  
 R:Matthews, L.  
 submitted to the EMBL Data Library, June 1996  
 A:Reference number: Z19240  
 A:Accession: T20227  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-93 <WIL>  
 A:Cross-references: EMBL:Z75532; PIDN:CAA99809.1; GSPDB:GN00023; CESP:C54G10.1  
 A:Experimental source: clone C54G10

C:Genetics:  
 A:Gene: CESP:C54G10.1  
 A:Map position: 5

Query Match 69.2%; Score 27; DB 2; Length 93;  
 Best Local Similarity 57.1%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EYVGSOH 7  
 DB 29 QVPGHH 35

# RESULT 6

D83599  
 conserved hypothetical protein PA0380 [imported] - Pseudomonas aeruginosa (strain PAO  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: D83599  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
 .; Lory, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; L  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: D83599  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-66 <STO>  
 A:Cross-references: GB:AE004475; GB:AE004091; NID:g9946221; PIDN:AA03769.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0380

Query Match 66.7%; Score 26; DB 2; Length 66;  
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPGSGH 7  
 DB 41 VPRSGH 46

# RESULT 7

Q06CW2  
 hypothetical protein A-73 trpr 5'-region - Escherichia coli (fragment)  
 C:Species: Escherichia coli  
 C:Date: 29-Sep-1999 #sequence\_revision 29-Sep-1999 #text\_change 29-Sep-1999  
 C:Accession: C93698; B93667; A04436  
 R:Singleton, C.K.; Roeder, W.D.; Bogosian, G.; Somerville, R.L.; Welch, H.L.  
 Nucleic Acids Res. 8, 1551-1560, 1980  
 A:Title: DNA sequence of the Escherichia coli trpr gene and prediction of the amino a  
 A:Reference number: A93698; MUID:81053831  
 A:Accession: C93698

A:Molecule type: DNA  
 A:Residues: 1-73 <STN>  
 A:Cross-references: GB:J01715; GB:V00369; NID:g148059  
 R:Gunsalus, R.P.; Yanofsky, C.  
 Proc. Natl. Acad. Sci. U.S.A. 77, 7117-7121, 1980  
 A:Title: Nucleotide sequence and expression of Escherichia coli trpr, the structural  
 A:Reference number: A93867; MUID:81175101  
 A:Accession: B93867  
 A:Molecule type: DNA  
 A:Residues: 1-33 <GUN>  
 A:Cross-references: GB:J01715; GB:V00369; NID:g148059  
 C:Comment: This is the hypothetical translation of a sequence that was not reported a  
 C:Genetics:  
 A:Map position: 100 min  
 A:Start codon: GTG

Query Match 66.7%; Score 26; DB 4; Length 73;  
 Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EPPGSOH 7  
| | | |  
Db 66 EPPGSKH 72

## RESULT 8

hypothetical protein PA2021 [imported] - Pseudomonas aeruginosa (strain PA01)  
G83393  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83393  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83393  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-74 <STO>  
A:Cross-references: GB:AE004628; GB:AE004091; NID:9948018; PIDN:AG05409.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2021

Query Match 66.7%; Score 26; DB 2; Length 74;  
Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EPPGSOH 7  
| | | |  
Db 15 EPPGSKH 21

## RESULT 9

hypothetical protein a376R - Chlorella virus PBCV-1  
T17876  
C:Species: Chlorella virus PBCV-1  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17876  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T17876  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-84 <GRA>  
A:Cross-references: EMBL:U042580; NID:84028896; PIDN:AAC96744.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Note: a376R

Query Match 66.7%; Score 26; DB 2; Length 84;  
Best Local Similarity 66.7%; Pred. No. 2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPGSOH 7  
| | | |  
Db 60 LPGSKH 65

## RESULT 10

hypothetical protein BH2324 [imported] - Bacillus halodurans (strain C-125)  
D83940  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: D83940  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83940  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-84 <STO>  
A:Cross-references: GB:AP001515; GB:BA000004; NID:910174886; PIDN:BA06043.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2324

Query Match 66.7%; Score 26; DB 2; Length 84;  
Best Local Similarity 50.0%; Pred. No. 2e+02;  
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSOH 7  
| | | |  
Db 50 IPGAKH 55

## RESULT 11

attachment protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid  
AF3175  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AF3175  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moore, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClure, K.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AF3175  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-96 <KUR>  
A:Cross-references: GB:AE008687; PIDN:AL45820.1; PID:917743559; GSPDB:GN00188  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: attD  
A:Genome: plasmid

Query Match 66.7%; Score 26; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSO 6  
| | | |  
Db 90 VPGSO 94

## RESULT 12

hypothetical protein RV2561 - Mycobacterium tuberculosis (strain H37Rv)  
E70728  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: E70728  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, R.; Janczewska, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: E70728  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-97 <COL>  
A:Cross-references: GB:Z77250; GB:AL123456; NID:93261617; PIDN:CAB01043.1; PID:e25531  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2561

Query Match 66.7%; Score 26; DB 2; Length 97;  
 Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 VPGSQH 7  
 :|||  
 Db 63 PGGQH 67

## RESULT 13

JY0104

hypothetical 11.4k protein (motA 3' region) - phage T4

C:Species: phage T4

A:Note: host Escherichia coli

C&gt;Date: 23-Oct-1992 #sequence\_revision 23-Oct-1992 #text\_change 08-Oct-1999

C:Accession: JY0104

R:Uzan, M.; Brody, E.; Favre, R.

Mol. Microbiol. 4, 1487-1496, 1990

A&gt;Title: Nucleotide sequence and control of transcription of the bacteriophage T4 motA

A:Reference number: JY0101; MUID:91141300

A:Accession: JY0104

A:Molecule type: DNA

A:Residues: 1-99 &lt;UZA&gt;

A:Cross-references: GB:Z48569; NID:9695449; PIDN:CAA8456.1; PID:9695453

C:Genetics:

A:Gene: motA-3

## Query Match

Best Local Similarity 50.0%; Score 26; DB 2; Length 99;  
 Pred. No. 2.3e+02;

Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
 :|||:|

Db 27 IPGARH 32

## RESULT 14

A39328

notechis II-5b nontoxic venom protein - common tiger snake (fragment)

C:Species: Notechis scutatus scutatus (common tiger snake, mainland tiger snake)

C&gt;Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 30-Sep-1993

C:Accession: A39328

R:Yang, C.C.; Chang, L.S.; Wu, F.S.

Toxicol. 29, 1337-1344, 1991

A&gt;Title: Venom constituents of Notechis scutatus scutatus (Australian tiger snake) from

A:Reference number: A39328; MUID:92263371

A:Accession: A39328

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-20 &lt;YAN&gt;

C:Superfamily: phospholipase A2

## Query Match

Best Local Similarity 64.1%; Score 25; DB 2; Length 20;  
 Pred. No. 69;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQ 6  
 :|||

Db 13 IPGSO 17

## RESULT 15

A60222

myelin basic protein - mouse (fragment)

C:Species: Mus musculus (house mouse)

C&gt;Date: 21-Oct-1992 #sequence\_revision 21-Oct-1992 #text\_change 17-Jul-1994

C:Accession: A60222

R:Aruga, J.; Okano, H.; Mikoshiba, K.

J. Neurochem. 56, 1222-1226, 1991

A&gt;Title: Identification of the new isoforms of mouse myelin basic protein: the existence

A:Reference number: A60222; MUID:91162193

A:Accession: A60222

A:Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-33 <AKU>

## Query Match

Best Local Similarity 64.1%; Score 25; DB 2; Length 33;  
 Pred. No. 1.2e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
 :|||  
 Db 3 VPGDHH 8Search completed: October 24, 2002, 15:37:41  
 Job time: 8.27869 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:23:28 : Search time 2.52459 Seconds  
(without alignments)  
107.359 Million cell updates/sec

Title: US-09-786-648-2

Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 12261

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	69.2	87	1	TAT_HV1MA
2	25	64.1	81	1	HPIS_TH1PF
3	24	61.5	61	1	NXSL_BOJAN
4	24	61.5	73	1	VF09_0REN2
5	24	61.5	74	1	VF09_0REN7
6	24	61.5	86	1	TRX1_HUMAN
7	24	61.5	88	1	FXV3_MOUSE
8	24	61.5	88	1	FXV3_PIG
9	24	61.5	91	1	Y13F_BP74
10	23.5	60.3	92	1	GONL_CHICK
11	23	59.0	34	1	HEMO_CHICK
12	23	59.0	55	1	RLX_HALMA
13	23	59.0	59	1	EA57_HUMAN
14	23	59.0	64	1	VGLE_HSVB
15	23	59.0	83	1	PGCA_PIG
16	23	59.0	87	1	Y14B_BP74
17	23	59.0	88	1	LCRS_YERPE
18	23	59.0	88	1	RS12_MYCS2
19	23	59.0	90	1	VAF1_DROPS
20	23	59.0	92	1	Z600_DROME
21	23	59.0	92	1	EM6_ARATH
22	22	56.4	14	1	PK6_PERAM
23	22	56.4	27	1	GRP_CANFA
24	22	56.4	36	1	CBA_NITVU
25	22	56.4	50	1	PRVM_SCYCA
26	22	56.4	58	1	TAT_HV1B5
27	22	56.4	61	1	AERC_AERSO
28	22	56.4	61	1	R14A_MYCTU
29	22	56.4	61	1	RS14_MYCLE
30	22	56.4	64	1	Q926F8_MYCDBACTE
31	22	56.4	64	1	Q926F8_CHLAMYDIA
32	22	56.4	64	1	RL35_CHLTPN
33	22	56.4	64	1	SCX2_MESMA
			64	1	SCX3_LEIOU

34	22	56.4	64	1	SCX4_MESMA	P58328 mesobuthus
35	22	56.4	67	1	CSPA_BACCE	O45096 bacillus ce
36	22	56.4	69	1	ATSL_AMACA	P80211 amarantulus
37	22	56.4	69	1	PEV_LOPAM	P09475 lophius ame
38	22	56.4	72	1	Y738_SYNV3	P74794 synechocyst
39	22	56.4	75	1	DISC_TRIEL	P23323 trimeresuru
40	22	56.4	75	1	YCEK_ECOLI	P45806 escherichia
41	22	56.4	79	1	SCX9_MESMA	P45698 mesobuthus
42	22	56.4	81	1	YTIU_ECOLI	P32164 escherichia
43	22	56.4	83	1	DIS1_BITAR	P17497 bitis arlet
44	22	56.4	84	1	SC13_MESMA	O9njc8 mesobuthus
45	22	56.4	84	1	SC16_MESMA	O9gyv6 mesobuthus

## ALIGNMENTS

RESULT 1						
TAT_HV1MA	STANDARD:	PRT:	87 AA.			
AC P04613;						
DT 13-AUG-1987 (Rel. 05, Created)						
DT 13-AUG-1987 (Rel. 05, Last sequence update)						
DT 16-OCT-2001 (Rel. 40, Last annotation update)						
DE TAR protein (Transactivating regulatory protein).						
GN TAR.						
OS Human immunodeficiency virus type 1 (MAL isolate) (HIV-1).						
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.						
ON NCBI_TaxID=11697;						
RN [1]						
RP SEQUENCE FROM N.A.						
RX MEDLINE=86245056; Pubmed=2424612;						
RA Allison M., Main-Hobson S., Montagnier L., Sonigo P.;						
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis						
of two isolates from African patients.";						
RL Cell 46:63-74(1986).						
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE						
TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND						
ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR						
PROMOTER.						
CC -1- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).						
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.						
CC -----						
CC This SWISS-PROT entry is copyright. It is produced through a collaboration						
between the Swiss Institute of Bioinformatics and the EMBL outstation -						
the European Bioinformatics Institute. There are no restrictions on its						
use by non-profit institutions as long as its content is in no way						
modified and this statement is not removed. Usage by and for commercial						
entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
CC -----						
CC EMBL; X04415; CAA28015.1; -.						
DR HIV; K03456; TATSMAL.						
DR InterPro; IPR001831; HIV_Tat.						
DR Pfam; PF00539; Tat; 1.						
DR PRINTS; PR00055; HIVTATDOMAIN.						
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;						
KW AIDS.						
SO SEQUENCE 87 AA; 10036 MW; 3832412849D5BICE CRC64;						
Query Match 69.2% Score 27; DB 1; Length 87;						
Best Local Similarity 71.4% Pred. No. 53;						
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;						
QY 1 EVPGSQH 7						
1 1 1 1 1						
DB 71 EQPSSQH 77						
RESULT 2.						
HPIS_TH1PF	STANDARD:	PRT:	81 AA.			
ID HPIS_TH1PF						
AC P00263;						

```

DR 21-JUL-1986 (Rel. 01, Created)
DR 21-JUL-1986 (Rel. 01, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE High potential iron-sulfur protein (HiPIP).
GN HiP.
OS Thiacapsa pfennigii.
OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
OC Thiococcus.
OX NCBI_TaxID=1057;
RN [1]
RP SEQUENCE.
RC STRAIN=KING 8816;
RX MEDLINE=74107423; PubMed=4814341;
RA Tedro S.M., Meyer T.E., Kamen M.D.;
RT "Primary structure of a high potential iron-sulfur protein from the
RT photosynthetic bacterium Thiacapsa pfennigii."
RT J. Biol. Chem. 249:1182-1188(1974).
CC -1- FUNCTION: SPECIFIC CLASS OF HIGH-REDOX-POTENTIAL 4FE-4S
CC FERREDOXINS. FUNCTIONS IN ANAEROBIC ELECTRON TRANSPORT IN MOST
CC PURPLE AND IN SOME OTHER PHOTOSYNTHETIC BACTERIA AND IN AT LEAST
CC ONE GENUS (PARACOCCLUS) OF HALOPHILIC, DENITRIFYING BACTERIA.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC (HIPIP) FAMILY.
DR PIR: A00266; IHTF.
DR HSP; P00260; IBOY.
DR InterPro: IPR000170; HiPot_Ironsulf.
DR Pfam: PF01355; HiPIP; 1.
DR PRINTS: PR00374; HiPIPERDOXIN.
DR PROSITE: PS00596; HiPIP; 1.
KM Electron transport; Iron-sulfur; 4Fe-4S.
FT METAL 43
FT METAL 43
FT METAL 46
FT METAL 59
FT METAL 59
FT METAL 73
FT METAL 73
SQ SEQUENCE 81 AA; 8951 MW; CE822A7B72598106 CRC64;

Query Match
Best Local Similarity 64.1%; Score 25; DB 1; Length 81;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPSQ 6
DB 35 ELPGSE 40

RESULT 3
NXSL_BOUN
ID NXSL_BOUN STANDARD; PRT; 61 AA.
AC P34075;
DR 01-FEB-1994 (Rel. 28, Created)
DR 01-FEB-1994 (Rel. 28, Last sequence update)
DR 01-FEB-1994 (Rel. 28, Last annotation update)
DE Short neurotoxin 1.
OS Boulengerina annulata annulata (Banded water cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubridae;
OC Elapidae; Bungarinae; Boulengerina.
OX NCBI_TaxID=8610;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=92263369; PubMed=1814007;
RA Weinstein S.A., Schmidt J.J., Smith L.A.;
RT "Lethal toxins and cross-neutralization of venoms from the African
RT water cobras, Boulengerina annulata annulata and Boulengerina
RT christyi."
RT Toxicon 29:1315-1327(1991).
RL -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
CC ACETYLCHOLINE RECEPTOR.
CC -1- MISCELLANEOUS: LD(50) IS 0.052 MG/KG BY INTRAPERITONEAL INJECTION.
DR PIR: A39327; A39327.

```

```

DR HSP; P01426; INEA.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00087; toxin; 1.
DR ProDom: PD000206; Snake_toxin; 1.
DR PROSITE: PS00272; SNAKE_TOXIN; 1.
KM Venom; Neurotoxin; postsynaptic neurotoxin.
FT DISULFID 3
FT DISULFID 23
FT DISULFID 17
FT DISULFID 40
FT DISULFID 42
FT DISULFID 53
FT DISULFID 54
FT DISULFID 59
SQ SEQUENCE 61 AA; 6837 MW; D329A0242FC32A3C CRC64;

Query Match
Best Local Similarity 61.5%; Score 24; DB 1; Length 61;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PGSOH 7
DB 7 PSSQH 11

RESULT 4
VF09_ORFN2
ID VF09_ORFN2 STANDARD; PRT; 73 AA.
AC P52586;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DR 15-DEC-1998 (Rel. 37, Last annotation update)
DE Protein F9 homolog (Fragment).
GN A3R.
OS Oryzavirus (strain NZ2) (OV NZ-2).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Parapoxviruses.
OC NCBI_TaxID=10259;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94076465; PubMed=8254780;
RA Lytle D.J., Fraser K.M., Fleming S.B., Mercer A.A., Robinson A.J.;
RT "Homologs of vascular endothelial growth factor are encoded by the
RT Poxvirus orf virus."
RT J. Virol. 68:84-92(1994).
RL -1- SIMILARITY: SOME TO VAC ORF PROTEIN F9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: S67520; AAB29219.1; -.
CC DR NON_TER 1
CC FT SEQUENCE 73 AA; 7623 MW; CAD9FE74576A9E6E CRC64;

Query Match
Best Local Similarity 61.5%; Score 24; DB 1; Length 73;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPQSOH 7
DB 20 VPSRSH 25

RESULT 5
VF09_ORFN7
ID VF09_ORFN7 STANDARD; PRT; 74 AA.
AC P52587;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DR 15-DEC-1998 (Rel. 37, Last annotation update)
DE Protein F9 homolog (Fragment).
GN A3R.

```



OS Off virus (strain NZ7) (OV NZ-7).  
 CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 CC Parapoxvirus.  
 OX NCBI\_TaxID=73495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94076465; PubMed=8254780;  
 RA Lytle D.J., Fraser K.M., Fleming S.B., Mercer A.A., Robinson A.J.;  
 RT "Homologs of vascular endothelial growth factor are encoded by the  
 RT Poxvirus off virus";  
 RL J. Virol. 66:84-92(1994).  
 CC -1- SIMILARITY: TO VAC ORF PROTEIN F9.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: S67522; AAB29222.1; -.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 74 AA; 7827 MW; B49FB9D04C560C70 CRC64;  
 QY 2 VPGSOH 7 61.5%; Score 24; DB 1; Length 74;  
 Db 20 VPASRH 25 Best Local Similarity 66.7%; Pred. No. 1.9e+02; Mismatches 1; Indels 0; Gaps 0;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 6  
 IRX1\_HUMAN STANDARD; PRT; 86 AA.  
 ID IRX1\_HUMAN  
 AC P78415;  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Iroquois-class homeodomain protein IRX-1 (Fragment).  
 GN IRX1.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lewis M.T., Strickland P.A., Ross S., Snyder C.J., Daniel C.W.;  
 RT "IRX, a new family of human homeobox genes from the breast.";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE TALE/IRO FAMILY OF HOMEOBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U90308; AAB50006.1; -.  
 DR MIM: 606197; -.  
 DR InterPro: IPR001356; Homeobox.  
 DR PROSITE: PS00027; HOMEOBOX\_1; PARTIAL.  
 DR PROSITE: PS00071; HOMEOBOX\_2; 1.  
 KW Homeobox; DNA-binding; Nuclear protein.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 37 >86 HOMEOBOX (TALE-TYPE).  
 FT NON\_TER 86 86  
 FT SIGNAL 1 20  
 SQ SEQUENCE 86 AA; 9626 MW; 21C263AE9450D7B6 CRC64;

Query Match 61.5%; Score 24; DB 1; Length 86;  
 Best Local Similarity 80.0%; Pred. No. 2.3e+02; Mismatches 1; Indels 0; Gaps 0;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 PGVGH 7 80.0%; Score 24; DB 1; Length 86;  
 Db 9 PGVGH 13  
 RESULT 7  
 FXY3\_MOUSE STANDARD; PRT; 88 AA.  
 ID FXY3\_MOUSE  
 AC 061835;  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE FXYD domain-containing ion transport regulator 3 precursor (Chloride  
 DE conductance inducer protein Mat-8) (Mammary tumor 8 kDa protein)  
 DE (Phospholemmann-like).  
 GN FXYD3 OR MAT8 OR PLML.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Breast;  
 RX MEDLINE=95060797; PubMed=7970700;  
 RA Morrison B.W., Leder P.;  
 RT "neu and ras initiate murine mammary tumors that share genetic  
 RT markers generally absent in c-myc and int-2-initiated tumors.";  
 RL Oncogene 9:3417-3426(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP EXPRESSION.  
 RX MEDLINE=9518184; PubMed=7836447;  
 RA Morrison B.W., Moorman J.R., Kowdley G.C., Kobayashi Y.M., Jones L.R.,  
 RA Leder P.;  
 RT "Mat-8, a novel phospholemmann-like protein expressed in human breast  
 RT tumors, induces a chloride conductance in xenopus oocytes.";  
 RL J. Biol. Chem. 270:2116-2182(1995).  
 CC -1- FUNCTION: INDUCES A HYPERPOLARIZATION-ACTIVATED CHLORIDE CURRENT  
 CC WHEN EXPRESSED IN XENOPUS OOCYTES. MAY BE A MODULATOR CAPABLE OF  
 CC ACTIVATING ENDOGENOUS OOCYTE CHANNELS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A SUBSET OF MURINE BREAST TUMORS.  
 CC -1- MISCELLANEOUS: MARKER OF A CELL TYPE PREFERENTIALLY TRANSFORMED BY  
 CC NEU OR RAS ONCOPROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE FXYD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X93038; CAA63606.1; -.  
 DR EMBL: BC002039; AAH02039.1; -.  
 DR MGD: MGI:107497; Fxyd3.  
 DR InterPro: IPR000272; ATP1G1\_P1M\_MAT8.  
 DR Pfam: PF02038; ATP1G1\_P1M\_MAT8; 1.  
 DR PROSITE: PS01310; FXYD; 1.  
 KW Transmembrane; Signal; Ionic channel; Ion transport.  
 FT SIGNAL 1 20  
 FT SIGNAL 21 88  
 FT CHAIN 21 88  
 FT REGULATOR 3.

FT DOMAIN 21 38 EXTRACELLULAR (POTENTIAL).  
 FT TRAMSEM 39 59 POTENTIAL.  
 FT DOMAIN 60 88 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 88 AA; 9526 MW; 9CD61684B856E35D CRC64;

Query Match  
 Best Local Similarity 61.5%; Score 24; DB 1; Length 88;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGSQH 7  
 |||  
 DB 82 PGSAH 86

RESULT 8  
 FXY3\_PIG STANDARD; PRT; 88 AA.  
 AC 09797;  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE FXYD domain-containing ion transport regulator 3 precursor (Chloride  
 DE conductance inducer protein Mat-8) (Mammary tumor 8 kDa protein).  
 GN FXYD3 OR MAT8.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Gastric mucosa;  
 RX MEDLINE=99140476; PubMed=10206733;  
 RA Maeda M., Hamano K., Hirano Y., Suzuki M., Takahashi E.-I., Terada T.,  
 RA Futai M., Sato R.;  
 RT "Structures of P-type transporting ATPases and chromosomal locations  
 RT of their genes.";  
 RL Cell Struct. Funct. 23:315-323(1998).  
 CC -1- FUNCTION: INDUCES A HYPERPOLARIZATION-ACTIVATED CHLORIDE CURRENT  
 CC WHEN EXPRESSED IN XENOPUS OOCYTES. MAY BE A MODULATOR CAPABLE OF  
 CC ACTIVATING ENDOGENOUS OOCYTE CHANNELS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE FXYD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB015759; BAA35078.1; -  
 DR InterPro; IPR000272; ATP1G1\_PLM\_MAT8.  
 DR Pfam; PF02038; ATP1G1\_PLM\_MAT8; 1.  
 DR PROSITE; PS01310; FXYD; 1.  
 KW Transmembrane; Signal; Ionic channel; Ion transport.  
 FT SIGNAL 1 17  
 FT CHAIN 18 88  
 FT POTENTIAL.  
 FT FXYD DOMAIN-CONTAINING ION TRANSPORT  
 FT REGULATOR 3.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 39 59  
 FT DOMAIN 60 88  
 FT CYTOPLASMIC (POTENTIAL).  
 FT SEQUENCE 88 AA; 9314 MW; 6CC7810B90512E5A CRC64;

Query Match  
 Best Local Similarity 61.5%; Score 24; DB 1; Length 88;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGSQH 7  
 |||  
 DB 82 PGSAH 86

RESULT 9  
 Y13F\_BPT4 STANDARD; PRT; 91 AA.  
 ID Y13F\_BPT4  
 AC P39459;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 10.1 kDa protein in cd-psst intergenic region.  
 GN Y13F OR CD.3.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like phages.  
 OX NCBI\_TaxID=10665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis.";  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF158101; AAD42549.1; -  
 DR Hypothetical protein.  
 KW SEQUENCE 91 AA; 10131 MW; B0010922E122A21F CRC64;

Query Match  
 Best Local Similarity 61.5%; Score 24; DB 1; Length 91;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGSQH 7  
 |||  
 DB 58 PGSAH 62

RESULT 10  
 GON1\_CHICK STANDARD; PRT; 92 AA.  
 ID GON1\_CHICK  
 AC P37042; P20407;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Progesterone-binding protein I precursor [Contains: Gonadotropin releasing  
 DE (luteinizing hormone releasing hormone I) (GnRH I) (LHRH I)  
 DE (luteinizing hormone releasing hormone I) (GnRH-associated peptide I).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WHITE LEGHORN;  
 RX MEDLINE=94059355; PubMed=7902095;  
 RA Dunn I.C., Chen Y., Hook C., Sharp P.J., Sang H.M.;  
 RT "Characterization of the chicken preprogonadotropin-releasing  
 RT hormone-I gene.";  
 RL J. Mol. Endocrinol. 11:19-29(1993).  
 RN [2]  
 RP SEQUENCE OF 24-33.  
 RC TISSUE=Hypothalamus;  
 RX MEDLINE=82265778; PubMed=7050119;  
 RA King J.A., Millar R.P.;  
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing  
 RT hormone. II. Isolation and characterization.";  
 RL J. Biol. Chem. 257:10729-10732(1982).  
 RN [3]

```

RP SEQUENCE OF 24-33.
RC TISSUE-Hypothalamus;
RA King J.A., Millar R.P.;
RT "Structure of avian hypothalamic gonadotrophin-releasing hormone.";
RL S. Afr. J. Sci. 78:124-125(1982).
[4]
RP SYNTHESIS OF 24-33.
RX MEDLINE-82265777; PubMed-7050118;
RA King J.A., Millar R.P.;
RT "Structure of chicken hypothalamic luteinizing hormone-releasing hormone. I. Structural determination on partially purified material.";
RL J. Biol. Chem. 257:10722-10728(1982).
CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE GnRH FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: X69491; CAA49246.1; -.
DR PIR: S33507; S33507.
DR InterPro: IPR002012; GnRH.
DR Pfam: PF00446; GnRH. 1.
DR PRINTS: PR01541; GONADOLIBRN1.
DR PROSITE: PS00473; GnRH. 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 92 PROGONADOLIBERIN I.
FT PEPTIDE 24 33 GONADOLIBERIN I.
FT PEPTIDE 37 92 GnRH-ASSOCIATED PEPTIDE I.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID.
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).
SQ SEQUENCE 92 AA; 10206 MW; 61AE87EBAF508B6A CRC64;

Query Match 60.3%; Score 23.5; DB 1; Length 92;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1 EVPGS-OH 7
Db 62 ECPGSYOH 69

RESULT 11
HEMO_CHICK STANDARD; PRT; 34 AA.
ID HEMO_CHICK
AC P20057;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hemopexin (Fragment).
DE Hemopexin (Fragment).
GN HPX.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE.
RX MEDLINE-88339942; PubMed-3421961;
RA Wellner D., Cheng K.C., Mueller-Eberhard U.;
RT "N-terminal amino acid sequences of the hemopexins from chicken, rat
RT and rabbit.";
RL Biochem. Biophys. Res. Commun. 155:622-625(1988).
CC -1- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN

```

```

CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE
CC CIRCULATION.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE LIVER, FOUND IN PLASMA.
DR PIR: C31514; C31514.
DR InterPro: IPR000585; Hemopexin.
DR PROSITE: PS00024; HEMOPEXIN; PARTIAL.
KW Glycoprotein; Heme; Plasma; Repeat; Acute phase; Transport.
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3598 MW; 07A2EAC27386C57 CRC64;

Query Match 59.0%; Score 23; DB 1; Length 34;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGSOH 7
Db 11 PGDEH 15

RESULT 12
RLX_HALMA STANDARD; PRT; 55 AA.
ID RLX_HALMA
AC P14125;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein Lx (HL32).
GN RPLX.
OS Haloarcula marismortui (Halo bacterium marismortui).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.
OX NCBI_TaxID=2238;
RN [1]
RP SEQUENCE.
RX MEDLINE-90076190; PubMed-2591382;
RA Hatakeyama T., Kaufmann F., Schroeter B., Hatakeyama T.;
RT "Primary structures of five ribosomal proteins from the
RT archaeobacterium Halobacterium marismortui and their structural
RT relationships to eubacterial and eukaryotic ribosomal proteins.";
RL Eur. J. Biochem. 185:685-693(1989).
CC -1- SIMILARITY: BELONGS TO THE LX FAMILY OF RIBOSOMAL PROTEINS.
DR PIR: S06848; RSHS32.
DR InterPro: IPR002768; Ribosomal_LX.
DR Pfam: PF01911; Ribosomal_LX; 1.
DR ProDom: PD010302; Ribosomal_LX; 1.
KW Ribosomal protein.
FT INIT_MET 0
SQ SEQUENCE 55 AA; 6257 MW; 7500AE44FC986529 CRC64;

Query Match 59.0%; Score 23; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GSOH 7
Db 39 GSQH 42

RESULT 13
EA57_HUMAN STANDARD; PRT; 59 AA.
ID EA57_HUMAN
AC O43247; O9Y4V8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE EA57 protein (Fragment).
GN EA57.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Fiedler M.:
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Lloyd D.:
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11142; CAA72029.1; -
DR EMBL; 282180; CAA45369.1; -
FT NON_TER
FT CONFLICT
SQ SEQUENCE 59 AA: 6471 MW: DFT539F9732CAF3 CRC64;

Query Match
Best Local Similarity 59.0%; Score 23; DB 1; Length 59;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 VPGSOH 7
DB 29 VPASHH 34

RESULT 14
VGLF_HSVB
ID VGLF_HSVB STANDARD; PRT; 64 AA.
AC P30816;
DR 01-JUL-1993 (Rel. 26, Created)
DR 01-JUL-1993 (Rel. 26, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein E homolog (Fragment).
GN GE.
OS Simian herpes B virus (Cercopithecid herpesvirus 1) (Shv).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10325;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92113572; PubMed=1309859;
RA Killeen A.M., Harrington L., Wall L.V.M., Kelly D.C.;
RT "Nucleotide sequence analysis of a homologue of herpes simplex virus
RT type 1 gene US9 found in the genome of simian herpes B virus.";
RL J. Gen. Virol. 73:195-199(1992).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN E FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S75996; AAB21001.1; -
DR PIR; P00277; P00277.
KW Glycoprotein.
FT NON_TER
FT CONFLICT
SQ SEQUENCE 64 AA: 7106 MW: 73748A1925C08CA3 CRC64;

Query Match
Best Local Similarity 59.0%; Score 23; DB 1; Length 64;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQ 6
DB 34 EAPGSR 39

```

```

RESULT 15
PCGA_PIG
ID PCGA_PIG STANDARD; PRT; 83 AA.
AC 029011;
DR 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DR 01-NOV-1997 (Rel. 35, Last annotation update)
DE Aggrecan core protein (cartilage-specific proteoglycan core protein)
DE (CSPCP) (Fragment).
GN AGC1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=cartilage; PubMed=7827755;
RX MEDLINE=95128522;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian
RT aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S74664; AAC60528.1; -
DR InterPro; IPR000538; Link.
DR InterPro; IPR001304; Lectin_C.
DR PROSITE; PS01241; LINK; PARTIAL.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; PARTIAL.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; PARTIAL.
KW Glycoprotein; Cartilage; Proteoglycan; Repeat.
FT NON_TER
FT DOMAIN
FT NON_TER
FT CONFLICT
SQ SEQUENCE 83 AA: 8576 MW: 99DC5641CBDC44A8 CRC64;

Query Match
Best Local Similarity 59.0%; Score 23; DB 1; Length 83;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGS 5
DB 70 ELPGS 74

```

Search completed: October 24, 2002, 15:35:29  
Job time : 4.52459 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:24:04 ; Search time 8.14754 Seconds  
(without alignments) 148.629 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EYFGSQH 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

107123

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.potent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	32	82.1	95	10	Q9FQJ8
2	31	79.5	95	10	Q9FQJ8
3	30	76.9	65	10	Q9M4H3
4	29	74.4	93	2	Q9RIJ9
5	28	71.8	58	4	Q9S1A6
6	28	71.8	71	3	Q9P8B1
7	27	69.2	31	6	Q292A1
8	27	69.2	47	12	Q9IL16
9	27	69.2	59	16	Q984D1
10	27	69.2	61	5	Q77288
11	27	69.2	69	4	Q9BU94
12	27	69.2	77	17	Q97C07
13	27	69.2	87	4	Q15962
14	27	69.2	87	10	Q81376
15	27	69.2	87	10	Q9FV06
16	27	69.2	87	10	Q9FV06

17	27	69.2	89	10	Q81995	Q81995 lycopersico
18	27	69.2	93	5	Q9U194	Q9U194 leishmania
19	27	69.2	93	5	Q18842	Q18842 caenorhabdit
20	26	66.7	50	12	Q9QT23	Q9QT23 tt virus. o
21	26	66.7	52	6	Q9T082	Q9T082 canis famli
22	26	66.7	63	7	Q9TPH2	Q9TPH2 mastacembi
23	26	66.7	64	11	Q92022	Q92022 rattus norv
24	26	66.7	66	16	Q916B5	Q916B5 pseudomonas
25	26	66.7	74	16	Q91293	Q91293 pseudomonas
26	26	66.7	77	11	Q9Q108	Q9Q108 rattus norv
27	26	66.7	84	12	Q98428	Q98428 paramecium
28	26	66.7	84	16	Q9KAG3	Q9KAG3 bacillus ha
29	26	66.7	94	10	Q9CAG6	Q9CAG6 arabidopsis
30	26	66.7	96	2	Q32733	Q32733 agrobacteri
31	26	66.7	97	4	Q9BRU2	Q9BRU2 homo sapien
32	26	66.7	99	8	Q37764	Q37764 mytilus edu
33	25	64.1	26	15	Q902V2	Q902V2 simian t-ce
34	25	64.1	26	15	Q902V1	Q902V1 simian t-ce
35	25	64.1	26	15	Q902V0	Q902V0 simian t-ce
36	25	64.1	27	13	Q90WP2	Q90WP2 fugu rubrip
37	25	64.1	63	2	Q929S3	Q929S3 bacillus ha
38	25	64.1	64	4	Q16505	Q16505 homo sapien
39	25	64.1	67	11	Q90970	Q90970 rattus norv
40	25	64.1	67	16	Q98PD5	Q98PD5 rhizobium l
41	25	64.1	70	2	Q9KXV1	Q9KXV1 streptomyce
42	25	64.1	76	5	Q9XXE2	Q9XXE2 caenorhabdi
43	25	64.1	80	10	Q9ZT55	Q9ZT55 zea mays (m
44	25	64.1	83	5	Q9VQW3	Q9VQW3 drosophila
45	25	64.1	88	10	Q9XGV0	Q9XGV0 orobanche r

# ALIGNMENTS

RESULT 1	Q9FQJ8	PRELIMINARY:	PRT:	95 AA.
ID	Q9FQJ8			
AC	Q9FQJ8:			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).			
OS	Amaranthus quitensis.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Caryophyllidae; Caryophyllales; Amaranthaceae; Amaranthus.			
OX	NCBI_Taxid=107609;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TRANSPONON-COPIA-LIKE RETROTRANSPONON;			
RA	He Y., Sun M.;			
RT	"Reverse transcriptase sequence evolution in copia-like			
RT	retrotransposons in plants.";			
RL	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF239933; AAC44333.1; -			
DR	InterPro: IPR03006; IG_MHC.			
DR	PROSITE: PS00290; IG_MHC; UNKNOWN_1.			
FT	NON_TER			
FT	NON_TER			
FT	NON_TER			
SQ	SEQUENCE	95 AA;	11392 MW;	8E931447E1683C5C CRC64;
Query Match				
Best Local Similarity		82.1%;	Score 32;	DB 10;
Matches	5;	Conservative	71.4%;	Pred. No. 17;
			1;	Indels
				0;
				Gaps
				0;
QY	1 EYFGSQH 7			
DB	26 EYFGKEH 32			
RESULT 2	Q9FQJ8	PRELIMINARY:	PRT:	95 AA.
ID	Q9FQJ8			

AC 09F0B8;  
 DT 01-MAR-2001 (TEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).  
 OS Amaranthus quitensis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Amaranthaceae; Amaranthus.  
 OX NCBI\_TaxID=107609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPON-COPIA-LIKE RETROTRANSPOSON;  
 RA He Y., Sun M.;  
 RT "Reverse transcriptase sequence evolution in copia-like  
 RT retrotransposons in plants."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF232981; AAG44323.1; -  
 DR InterPro: IPR003006; IG\_MHC.  
 DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.  
 FT NON\_TER 1  
 FT 95  
 SQ SEQUENCE 95 AA; 11241 MW; 36403CBDC40D769A CRC64;

Query Match 79.5%; Score 31; DB 10; Length 95;  
 Best Local Similarity 57.1%; Pred. No. 28;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 1:11:11  
 DB 26 EIRGEH 32

RESULT 3  
 O9M4H3 PRELIMINARY; PRT; 65 AA.  
 AC 09M4H3;  
 DT 01-OCT-2000 (TEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)  
 DE PUTATIVE METALLOTHIONEIN-LIKE PROTEIN.  
 GN GRIP24.  
 OS Vitis vinifera (Grape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;  
 OC Vitis.  
 OX NCBI\_TaxID=29760;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. SHIRAZ; TISSUE-FRUIT;  
 RX MEDLINE=20177861; PubMed=10712544;  
 RA Davies C., Robinson S.P.;  
 RT "Differential screening indicates a dramatic change in mRNA profiles  
 RT during grape berry ripening. Cloning and characterization of cDNAs  
 RT encoding putative cell wall and stress response proteins."  
 RL Plant Physiol. 122:803-812(2000).  
 DR EMBL; AJ237990; CAB85630.1; -  
 SQ SEQUENCE 65 AA; 6777 MW; B5EA7D8D7B9170D4 CRC64;

Query Match 76.9%; Score 30; DB 10; Length 65;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 1:11:11  
 DB 40 EYPAQH 46

RESULT 4  
 O9RJ49 PRELIMINARY; PRT; 93 AA.  
 ID 09RJ49;  
 AC 09RJ49;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)  
 DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.  
 GN SC18.04C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL132644; CAB59435.1; -  
 SQ SEQUENCE 93 AA; 9977 MW; 0F65FBC323F57E0C CRC64;

Query Match 74.4%; Score 29; DB 2; Length 93;  
 Best Local Similarity 66.7%; Pred. No. 74;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
 :11:11  
 DB 21 IPGQH 26

RESULT 5  
 O95146 PRELIMINARY; PRT; 58 AA.  
 AC 095146;  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
 DE LAGE-2ALF PROTEIN ISOFORM.  
 GN LAGE-2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.-F., Johnston S.L., Zeng G., Topalian S.L.,  
 RA Schwartzentruber D.J., Rosenberg S.A.;  
 RT "A Breast and Melanoma-Shared Tumor Antigen: T Cell Responses to  
 RT Antigenic Peptides Translated from Different Open Reading Frames."  
 RL J. Immunol. 161:3596-3606(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lethe B.G.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99454989; PubMed=10523621;  
 RA De Smet C., Lurquin C., Lethe B., Martelange V., Boon T.;  
 RT "DNA methylation is the primary silencing mechanism for a set of germ  
 RT line- and tumor-specific genes with a CpG-rich promoter."  
 RL Mol. Cell. Biol. 19:7327-7335(1999).  
 DR EMBL; AF038567; AAD05203.1; -  
 DR EMBL; AJ275977; CAB76944.1; -  
 SQ SEQUENCE 58 AA; 6188 MW; ED12057564BC7EF2 CRC64;

Query Match 71.8%; Score 28; DB 4; Length 58;  
 Best Local Similarity 83.3%; Pred. No. 74;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
 ||||:|  
 DB 30 EVPGAQ 35

## RESULT 6

Q9P8B1 PRELIMINARY; PRT; 71 AA.  
 AC Q9P8B1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE METALLOTHIONEIN (FRAGMENT).  
 GN MT2.  
 OS Agaricus bisporus (Common mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Agaricaceae; Agaricus.  
 OX NCBI\_TaxID=5341;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HORST U3;  
 RA Eastwood D.C., Kingsnorth C.S., Jones H., Burton K.S.;  
 RT "Development of Agaricus bisporus sporophores following tissue  
 RT detachment is maintained by the expression of stress tolerance and  
 RT nutritional genes."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ271695; CAB5689.1; -;  
 DR InterPro: IPR002086; Aldehyde\_dehydr.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CTS; UNKNOWN\_1.  
 FT NON\_TER 1 1  
 FT NON\_TER 71 71  
 SQ SEQUENCE 71 AA; 7311 MW; 8CE9E663110CEA3 CRC64;

Query Match 71.8%; Score 28; DB 3; Length 71;  
 Best Local Similarity 71.4%; Pred. No. 92;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSQ 7  
 ||||:|  
 DB 44 EVPKNOH 50

## RESULT 7

Q29241 PRELIMINARY; PRT; 31 AA.  
 AC Q29241;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ORF PROTEIN (FRAGMENT).  
 GN ORF.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SMALL INTESTINE;  
 RX MEDLINE=96327607; PubMed=8672129;  
 RA "Walter A.R., Fredholm M., Davies W.;  
 RT "Evaluation and characterization of a porcine small intestine CDNA  
 RT library".  
 RL Mamm. Genome 7:509-517(1996).  
 DR EMBL: F15070; CAA23326.1; -;  
 FT NON\_TER 1 1  
 FT NON\_TER 31 31  
 SQ SEQUENCE 31 AA; 3376 MW; 49CF21FECAB1613 CRC64;

Query Match 69.2%; Score 27; DB 6; Length 31;  
 Best Local Similarity 66.7%; Pred. No. 63;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSOH 7  
 ||||:|  
 DB 11 VPGARH 16

## RESULT 8

Q96TC5 PRELIMINARY; PRT; 32 AA.  
 AC Q96TC5;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DEOXYRIBONUCLEASE II (FRAGMENT).  
 GN DNASE2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99123683; PubMed=9924608;  
 RA Yasuda T., Takeshita H., Iida R., Tsutsumi S., Nakajima T., Hosomi O.,  
 RA Nakashima Y., Mori S., Kishi K.;  
 RT "Structure and organization of the human deoxyribonuclease II (DNase  
 RT II) gene."  
 RL Ann. Hum. Genet. 62:299-305(1998).  
 DR EMBL: AB008565; BAB5597.1; -;  
 FT NON\_TER 32 32  
 SQ SEQUENCE 32 AA; 3195 MW; AC245182DC696790 CRC64;

Query Match 69.2%; Score 27; DB 4; Length 32;  
 Best Local Similarity 66.7%; Pred. No. 65;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
 ||||:|  
 DB 14 QIPGSO 19

## RESULT 9

Q91L16 PRELIMINARY; PRT; 47 AA.  
 AC Q91L16;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE FUSION PROTEIN (FRAGMENT).  
 GN F.  
 OS Human metapneumovirus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Metapneumovirus.  
 OX NCBI\_TaxID=162145;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=93-1;  
 RX MEDLINE=21279106; PubMed=11385510;  
 RA van den Hoogen B.G., de Jong J.C., Groen J., Kuiken T., de Groot R.,  
 RA Fouchier R.A., Osterhaus A.D.;  
 RT "A newly discovered human pneumovirus isolated from young children  
 RT with respiratory tract disease".  
 RL Nat. Med. 7:719-724(2001).  
 DR EMBL: AF371341; AAK62945.1; -;  
 FT NON\_TER 1 1  
 FT NON\_TER 47 47  
 SQ SEQUENCE 47 AA; 5142 MW; 15A0DAA7C7C472BF CRC64;

Query Match 69.2%; Score 27; DB 12; Length 47;  
 Best Local Similarity 71.4%; Pred. No. 98;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EYVGSOH 7  
11111  
DB 40 EYEGEOH 46

## RESULT 10

Q984D1 PRELIMINARY; PRT; 59 AA.  
AC Q984D1;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE MSR8054. PROTEIN.  
GN MSR8054.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
Mochizuki C., Nakayama S., Nakazaki N., Shingo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003013; BAB53699.1; -.  
KW Complete proteome.  
SQ SEQUENCE 59 AA; 6023 MW; 0F330D588FA94FB5 CRC64;

Query Match 69.2%; Score 27; DB 16; Length 59;  
Best Local Similarity 80.0%; Pred. No. 1.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 PGSOH 7  
11111  
DB 8 PGAOH 12

## RESULT 11

077288 PRELIMINARY; PRT; 61 AA.  
AC 077288;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PROS26 PROTEIN (FRAGMENT).  
GN PROS26 OR CG4097.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Boukharem N., Bichon A., Gay P., Dru P., Petitjean A.M., Lemeunier F.,  
RA Contamine D.;  
RT "Genetic and molecular features of Su(P), a gene that interacts with  
ref(2)P in male fertility of Drosophila melanogaster";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ011320; CAA09596.1; -.  
DR HSSD; P23724; ITRP.  
DR FlyBase; FBgn002284; Pros26.  
DR InterPro; IPR001353; Proteasome.  
DR Pfam; PF00227; proteasome; 1.  
FT NON\_TER 61  
SQ SEQUENCE 61 AA; 6706 MW; 1179A3032EA6486E CRC64;

Query Match 69.2%; Score 27; DB 5; Length 61;  
Best Local Similarity 57.1%; Pred. No. 1.3e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSOH 7  
11111  
DB 13 QYVGMKH 19

## RESULT 12

Q9BU94 PRELIMINARY; PRT; 69 AA.  
AC Q9BU94;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SIMILAR TO HYPOTHETICAL PROTEIN FLJ20378.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PLACENTA, CHORIOCARCINOMA;  
RA Strusberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC002811; AAH02811.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 69 AA; 7649 MW; 48BA04F9785E8098 CRC64;

Query Match 69.2%; Score 27; DB 4; Length 69;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
11111  
DB 39 VPGRSH 44

## RESULT 13

097C07 PRELIMINARY; PRT; 77 AA.  
AC 097C07;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE LEUCINE-RESPONSIVE REGULATORY PROTEIN.  
GN TVG0307586.  
OS Thermoplasma volcanium.  
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
OC Thermoplasma.  
OX NCBI\_TaxID=50339;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GSSI / DSM 4299 / JCM 9571;  
RX MEDLINE=20570466; PubMed=11121031;  
RA Kawashima T., Ameno N., Koike H., Makino S.-I., Higuchi S.,  
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanenori K., Kawamoto T.,  
RA Nunoshita T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;  
RT "Archaeal adaptation to higher temperatures revealed by genomic  
sequence of Thermoplasma volcanium";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).  
DR EMBL; AP000992; BAB59440.1; -.  
KW Complete proteome.  
SQ SEQUENCE 77 AA; 8586 MW; E5BC84D7CAE7E464 CRC64;

Query Match 69.2%; Score 27; DB 17; Length 77;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
11111  
DB 11 VPGRSH 16



Search completed: October 24, 2002, 15:36:49  
Job time : 11.1475 secs

# RESULT 14

015962 PRELIMINARY; PRT; 87 AA.  
AC 015962;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE THYROTROPIN BETA SUBUNIT (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCB1\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93103017; PubMed-8416675;  
RA Peele M.E., Carr F.E., Baker J.R.Jr., Wartofsky L., Burman K.D.;  
RT "TSH beta subunit gene expression in human lymphocytes.";  
RL Am. J. Med. Sci. 305:1-7(1993).  
DR EMBL: S51112; AAB24571.2; -.  
DR HSSP: P01233; 1XUL.  
DR InterPro: IPR000359; Cys\_knot.  
DR InterPro: IPR001545; Glyco\_hormone\_beta.  
DR Pfam: PF00007; Cys\_knot; 1.  
DR SMART: SM00068; GHB; 1.  
DR PROSITE: PS00689; GLYCO\_HORMONE\_BETA\_2; 1.  
FT NON\_TER 1  
FT NON\_TER 87  
SQ SEQUENCE 87 AA; 13AEEEF7F9728D930 CRC64;

## Query Match

Best Local Similarity 69.2%; Score 27; DB 4; Length 87;  
Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
1:1:1  
Db 35 EIPGCPH 41

## RESULT 15

081376 PRELIMINARY; PRT; 87 AA.  
ID 081376;  
AC 081376;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE REVERSE TRANSCRIPTASE (FRAGMENT).  
GN RT.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCB1\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. H722; TISSUE=LEAF;  
RA Rogers S.A., Pauls K.P.;  
RT "Y1-copia class retrotransposons of tomato."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF072638; AAC34605.1; -.  
FT NON\_TER 1  
FT NON\_TER 87  
SQ SEQUENCE 87 AA; 10302 MW; 2441ECF4E1456FF6 CRC64;

## Query Match

Best Local Similarity 69.2%; Score 27; DB 10; Length 87;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
1:1:1  
Db 22 VPGKEH 27

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:17 ; Search time 11.2459 Seconds  
(without alignments)  
69.138 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EYVGSQH 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 10%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	7	21	AA87460
2	39	100.0	8	21	AA87464
3	39	100.0	12	21	AA87461
4	39	100.0	15	10	AA87461
5	39	100.0	15	16	AA87462
6	39	100.0	21	21	AA87463
7	39	100.0	21	21	AA87463
8	39	100.0	23	16	AA87463
9	39	100.0	26	4	AA87463
10	39	100.0	41	6	AA87463
11	39	100.0	46	6	AA87463

12	39	100.0	47	4	AA87460
13	39	100.0	93	16	AA87464
14	39	100.0	93	20	AA87461
15	39	100.0	93	20	AA87461
16	39	100.0	93	21	AA87461
17	39	100.0	93	21	AA87461
18	39	100.0	103	6	AA87461
19	39	100.0	103	17	AA87461
20	39	100.0	103	17	AA87461
21	39	100.0	103	17	AA87461
22	39	100.0	103	17	AA87461
23	39	100.0	103	19	AA87461
24	39	100.0	103	22	AA87461
25	39	100.0	103	22	AA87461
26	39	100.0	103	22	AA87461
27	39	100.0	103	22	AA87461
28	39	100.0	105	22	AA87461
29	39	100.0	118	11	AA87461
30	39	100.0	123	22	AA87461
31	39	100.0	123	22	AA87461
32	39	100.0	124	10	AA87461
33	39	100.0	124	17	AA87461
34	39	100.0	124	17	AA87461
35	39	100.0	124	21	AA87461
36	39	100.0	124	21	AA87461
37	39	100.0	124	22	AA87461
38	39	100.0	124	22	AA87461
39	39	100.0	124	22	AA87461
40	39	100.0	124	22	AA87461
41	39	100.0	124	22	AA87461
42	39	100.0	124	22	AA87461
43	39	100.0	124	22	AA87461
44	39	100.0	124	22	AA87461
45	39	100.0	124	22	AA87461

#### ALIGNMENTS

RESULT 1  
AA87460 standard; peptide: 7 AA.

AC AA87460:

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:2.

KW Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;  
KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;  
KW adjuvant; immune disorder; diarrhoea.

OS Vibrio cholerae.

OS Escherichia coli.

PN WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-GH02970.

PR 07-SEP-1998; 98GB-0019484.

PI (UYBR-) UNIV BRISTOL.

PI Williams NA, Hirst TR;  
WPI, 2000-256943/22.

DR Derivatives of Escherichia coli heat labile enterotoxins useful as  
PT immunomodulators and for treating diarrhoea and which do not bind the  
PT glycolipid receptor GM-1 -

XX Claim 1: Page 13: 62pp; English.  
 PS The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.

SO Sequence 7 AA;

Query Match 100.0%; Score 39; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 |||||  
 Db 1 EYVGSQH 7

RESULT 2

ID AAY87464 standard; peptide; 8 AA.

XX AAY87464;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived immunomodulatory peptide.

XX Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KM adjuvant; immune disorder; diarrhoea.

XX Vibrio cholerae.

OS Escherichia coli.

XX WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-CB02970.

PR 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
 PT immunomodulators and for treating diarrhoea and which do not bind the  
 PT glycolipid receptor GM-1 -  
 XX Example 5: Page 45; 62pp; English.

CC The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention (AAY87461-Y87463) are  
 CC fragments of the beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the  
 CC same effects as normal EtxB and CtxB subunits, except that they do not  
 CC bind or cross link GM-1. They may be used in medicine as an  
 CC immunomodulator or adjuvant. They may also be used as an inhibitor for  
 CC toxin-induced diarrhoea. Therefore, the peptides may be used in the  
 CC production of a composition for treating, preventing and/or modulating a  
 CC disease associated with an immune disorder and/or toxin-induced  
 CC diarrhoea. Sequences AAY87464-Y87465 represent peptides used in an  
 CC exemplification of the present invention to assess whether a peptide  
 CC corresponding to a portion of the beta-4- alpha-2 loop of EtxB and CtxB  
 CC has immunomodulatory effects. Peptide AAY87464 corresponds to residues  
 CC 51-58 of the EtxB/CtxB beta-4- alpha-2 loop, and peptide AAY87465 is a  
 CC randomly selected control peptide.

SO Sequence 8 AA;

Query Match 100.0%; Score 39; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 |||||  
 Db 1 EYVGSQH 7

RESULT 3

ID AAY87461 standard; peptide; 12 AA.

XX AAY87461;

DT 03-JUL-2000 (first entry)

DE Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:3.

XX Cholera toxin subunit B; CtxB; heat labile enterotoxin subunit B; EtxB;

KW beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;

KM adjuvant; immune disorder; diarrhoea.

XX Vibrio cholerae.

OS Escherichia coli.

XX WO200014114-A1.

PD 16-MAR-2000.

PF 07-SEP-1999; 99WO-CB02970.

PR 07-SEP-1998; 98GB-0019484.

XX (UYBR-) UNIV BRISTOL.

XX Williams NA, Hirst TR;

DR WPI; 2000-256943/22.

PT Derivatives of Escherichia coli heat labile enterotoxins useful as  
 PT immunomodulators and for treating diarrhoea and which do not bind the  
 PT glycolipid receptor GM-1 -  
 PS Disclosure: Page 15; 62pp; English.

XX The invention relates to peptide fragments of the Escherichia coli heat  
 CC labile enterotoxin (Etx) and its closely related homologue, cholera  
 CC toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous  
 CC GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are  
 CC composed of one A subunit and five identical B subunits. The A subunit  
 CC is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-  
 CC ribosyltransferase activity, while the B subunits (EtxB and CtxB)  
 CC facilitate the entry of subunit A into the host cell via the binding and  
 CC cross-linking of GM-1 receptors. Although GM-1 binding is responsible  
 CC for some of the effects of Etx and Ctx, it has been found that certain  
 CC effects of the toxins, such as immunomodulation, are not mediated  
 CC through GM-1 binding. The peptides of the invention are fragments of the  
 CC beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as  
 CC normal EtxB and CtxB subunits, except that they do not bind or cross link  
 CC GM-1. They may be used in medicine as an immunomodulator or adjuvant.  
 CC They may also be used as an inhibitor for toxin-induced diarrhoea.  
 CC Therefore, the peptides may be used in the production of a composition  
 CC for treating, preventing and/or modulating a disease associated with an  
 CC immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463  
 CC represent preferred peptides of the invention, AAY87460 being  
 CC particularly preferred.

XX Sequence 12 AA;

Query Match 100.0%; Score 39; DB 21; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Gaps 0;

Matches 7; Conservative 0; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 |||||  
 Db 2 EYPGSOH 8

RESULT 4  
 AAP93498 ID AAP93498 standard; protein; 15 AA.

XX AAP93498;

XX 03-MAY-1990 (first entry)

XX CTP3 epitope of the Cholera toxin B subunit.

XX CTP3 epitope of Cholera toxin B subunit; flagellin fusion protein;

XX vaccine; immunotherapy; ds;

XX WO8910967-A.

XX 16-NOV-1989.

XX 05-MAY-1989; 89WO-US01932.

XX 05-MAY-1988; 88US-0190570.

XX (PRAX-) PRAXIS BIOLOGICS INC.

XX (STRD ) LEYLAND STANDFORD JUNIOR UNIV.

XX Marjarian WR, Stocker BAD, Newton SMC;

XX WPI; 1989-356496/48.

XX N-PSDB; AAN92414.

XX New recombinant flagellin gene including sequence - for heterologous  
 PT epitope, and expressed fusion proteins, useful in vaccines and for prodn.  
 PT of antibodies.

XX Disclosure; fig.4B; 137pp; English.

XX This sequence corresponds to the CTP3 epitope of the Cholera toxin B  
 CC subunit. The DNA sequence encoding this ligates to othersynthetic  
 CC oligonucleotides to form a new recombinant gene. This encodes  
 CC a flagellin fusion protein which can be used in vaccines for immuno-

CC therapy.

XX Sequence 15 AA;

Query Match 100.0%; Score 39; DB 10; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.37; Mismatches 0; Gaps 0;

Matches 7; Conservative 0; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 |||||  
 Db 2 EYPGSOH 8

RESULT 5  
 AAR85125 ID AAR85125 standard; peptide; 15 AA.

XX AAR85125;

XX 13-JUN-1996 (first entry)

XX Cholera toxin B antigenic peptide fragment CTP3.

XX Conjugate; cholera; B toxin; peptide fragment; microparticulate;

XX inert carrier; modified silica; thyroglobulin; oral vaccine;

XX immunisation; infection; insoluble; digestive tract; antigen;

XX intestines; antibodies; secretory; Iga class.

XX Vibrio cholerae.

XX WO9529701-A1.

XX 02-MAY-1995; 95WO-EP01661.

XX 03-MAY-1994; 94IL-0109519.

XX (YEDA ) YEDA RES & DEV CO LTD.

XX Marks RS, Mirelman D, Sela M;

XX WPI; 1995-403805/51.

XX Vaccines for oral immunisation against infecting agents, e.g.

XX cholera - comprise a conjugate of an antigen of an infecting agent

XX covalently bound to micro:particulate inert carrier, e.g. modified

XX aldehyde silica

XX Claim 7; Page 25; 40pp; English.

XX A compsn. comprising a conjugate of an antigenic cholera B toxin

XX peptide fragment (e.g. AAR85126-30 and esp. AAR85125), covalently

XX bound to a microparticulate inert carrier (e.g. modified silica or

XX thyroglobulin) can be used as an oral vaccine for immunisation

XX against cholera infection. The inert carrier is insoluble in the

XX digestive tract, allowing presentation of the antigen in the

XX intestines, where it will elicit antibodies mainly of the

XX secretory Iga class.

XX Sequence 15 AA;

Query Match 100.0%; Score 39; DB 16; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.37; Mismatches 0; Gaps 0;

Matches 7; Conservative 0; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
 |||||  
 Db 2 EYPGSOH 8

RESULT 6  
 AAY87462

```

ID  AAY87462 standard; peptide: 21 AA.
XX
AC  AAY87462;
XX
DT  03-JUL-2000 (first entry)
DE  Cholera toxin B/enterotoxin B-derived peptide, SEQ ID NO:4.
XX
KW  Cholera toxin subunit B; CTxB; heat labile enterotoxin subunit B; EtxB;
KM  beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX  adjuvant; immune disorder; diarrhoea.
OS  Vibrio cholerae.
XX  Escherichia coli.
XX  WO200014114-A1.
XX
PD  16-MAR-2000.
XX
PF  07-SEP-1999; 99WO-GB02970.
XX
PR  07-SEP-1998; 98GB-0019484.
XX
PA  (UYBR-) UNIV BRISTOL.
XX
PI  Williams NA, Hirst TR;
XX  WPI; 2000-256943/22.
XX
DR  Derivatives of Escherichia coli heat labile enterotoxins useful as
PT  immunomodulators and for treating diarrhea and which do not bind the
XX  glycolipid receptor GM-1 -
XX
PS  Disclosure; Page 15; 62pp; English.
XX
CC  The invention relates to peptide fragments of the Escherichia coli heat
CC  labile enterotoxin (Etx) and its closely related homologue, cholera
CC  toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC  GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC  composed of one A subunit and five identical B subunits. The A subunit
CC  is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC  ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC  facilitate the entry of subunit A into the host cell via the binding and
CC  cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC  for some of the effects of Etx and Ctx, it has been found that certain
CC  effects of the toxins, such as immunomodulation, are not mediated
CC  through GM-1 binding. The peptides of the invention are fragments of the
CC  beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC  normal EtxB and CtxB subunits, except that they do not bind or cross link
CC  GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC  They may also be used as an inhibitor for toxin-induced diarrhoea.
CC  Therefore, the peptides may be used in the production of a composition
CC  for treating, preventing and/or modulating a disease associated with an
CC  immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC  represent preferred peptides of the invention, AAY87460 being
CC  particularly preferred.
XX
SQ  Sequence 21 AA:
XX
Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPGSOH 7
DB 7 EYPGSOH 13

```

```

XX
DT  03-JUL-2000 (first entry)
XX
DE  E. coli heat labile enterotoxin B-derived peptide, SEQ ID NO:5.
XX
KW  Heat labile enterotoxin subunit B; EtxB;
KM  beta-4-alpha-2 loop; GM-1 ganglioside receptor; immunomodulation;
XX  adjuvant; immune disorder; diarrhoea.
OS  Escherichia coli.
XX
PN  WO200014114-A1.
XX
PD  16-MAR-2000.
XX
PF  07-SEP-1999; 99WO-GB02970.
XX
PR  07-SEP-1998; 98GB-0019484.
XX
PA  (UYBR-) UNIV BRISTOL.
XX
PI  Williams NA, Hirst TR;
XX  WPI; 2000-256943/22.
XX
DR  Derivatives of Escherichia coli heat labile enterotoxins useful as
PT  immunomodulators and for treating diarrhea and which do not bind the
XX  glycolipid receptor GM-1 -
XX
PS  Disclosure; Page 15; 62pp; English.
XX
CC  The invention relates to peptide fragments of the Escherichia coli heat
CC  labile enterotoxin (Etx) and its closely related homologue, cholera
CC  toxin (Ctx) from Vibrio cholerae which do not bind to the ubiquitous
CC  GM-1 ganglioside receptors on host cell surfaces. Both Etx and Ctx are
CC  composed of one A subunit and five identical B subunits. The A subunit
CC  is responsible for toxicity, possessing adenosine diphosphate (ADP) ADP-
CC  ribosyltransferase activity, while the B subunits (EtxB and CtxB)
CC  facilitate the entry of subunit A into the host cell via the binding and
CC  cross-linking of GM-1 receptors. Although GM-1 binding is responsible
CC  for some of the effects of Etx and Ctx, it has been found that certain
CC  effects of the toxins, such as immunomodulation, are not mediated
CC  through GM-1 binding. The peptides of the invention are fragments of the
CC  beta-4- alpha-2 loop of EtxB and/or CtxB, exerting the same effects as
CC  normal EtxB and CtxB subunits, except that they do not bind or cross link
CC  GM-1. They may be used in medicine as an immunomodulator or adjuvant.
CC  They may also be used as an inhibitor for toxin-induced diarrhoea.
CC  Therefore, the peptides may be used in the production of a composition
CC  for treating, preventing and/or modulating a disease associated with an
CC  immune disorder and/or toxin-induced diarrhoea. Sequences AAY87460-Y87463
CC  represent preferred peptides of the invention, AAY87460 being
CC  particularly preferred.
XX
SQ  Sequence 21 AA:
XX
Query Match 100.0%; Score 39; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EYPGSOH 7
DB 7 EYPGSOH 13

```

```

RESULT 7
AAY87463
ID  AAY87463 standard; peptide: 21 AA.
XX
AC  AAY87463;

```

```

RESULT 8
AAR76748
ID  AAR76748 standard; Protein; 23 AA.
XX
AC  AAR76748;
XX
DT  18-MAR-1996 (first entry)
XX
DE  Residues 50-64 of cholera toxin B subunit and FimH 224-226.

```

```

XX  F1mH: type 1 fimbriae; organelle: adhesin; alpha-D-mannoside residue;
KM  F1mH: F1mH; receptor binding site; PCR; amplify; ss.
XX  Chimeric - Vibrio cholerae.
OS  Chimeric - Escherichia coli.
XX  Key
FH  MISC-difference 1 Location/Qualifiers
FT  Peptide /note= "Represents F1mH residue 224"
FT  Peptide /note= "Linker peptide"
FT  Peptide /note= "Cholera toxin B subunit 50-64"
FT  Peptide /note= "Linker peptide"
FT  MISC-difference 23 /note= "Represents F1mH residue 226"
XX  W09520657-A1.
XX  03-AUG-1995.
XX  27-JAN-1995; 95MO-DK00042.
XX  27-JAN-1994; 94US-0187166.
XX  (GXBI-) GX BIOSYSTEMS AS.
XX  Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
PI  WPI: 1995-275442/36.
DR  N-PSDB; AAO93061.
XX  Receptor specific bacterial adhesins - useful for targeting active
PT  compounds and microbial cells to locations of receptors
XX  Example 3; Page 58; 152pp; English.
XX  This sequence is encoded by a fragment of the the plasmid pLPA93
CC  which was used in the production of f1mH fusion genes comprising
CC  the cholera toxin B subunit inserted into the f1mH gene. This insert
CC  shows the inclusion of the B subunit into the f1mH protein at position
CC  224-226. The chimeric genes were then opt. further modified by insertion
CC  of the hepatitis B virus surface antigen pre-S2 region into a different
CC  position of the f1mH adhesin of type 1 fimbriae. Restriction site handles
CC  (Bg1II-sites) were introduced into the f1mH gene, and the foreign
CC  insertion of the epitopes did not significantly alter the adhesive
CC  on the surface of fimbriae on bacterial hosts illustrated the possibility
CC  of using bacterial adhesins as general presenters of foreign antigens and
CC  epitopes. These chimeric genes may be used in the production of variant
CC  f1mH adhesins which may be useful for targeting active compounds
CC  and microbial cells to locations comprising selected receptors to which
CC  the adhesins bind.
XX  Sequence 23 AA:
SQ
Query Match 100.0%; Score 39; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVPGSOH 7
DB 6 EVPGSOH 12

```

```

XX  21-APR-1992 (first entry)
DT  Sequence of amino acids 50-75 of the cholera toxin B1 subunit which
XX  carries an Arg at posns. 67 and 73.
DE  Cholera vaccine; therapy; E.coli infection; enterotoxin LT.
XX  Vibrio cholerae.
OS  EP95426-A.
XX  30-NOV-1983.
XX  26-MAY-1983; 83EP-0401052.
XX  26-MAY-1982; 82FR-0009167.
XX  (CNRS ) CNRS CENT NAT RECH SCI.
XX  (INSP ) INST PASTEUR.
XX  Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;
XX  Guyongruaz A, Delmas A;
DR  WPI: 1983-834645/49.
XX  Cholera toxin B, sub-unit polypeptide(s) as vaccines and
PT  medicaments - effective against Escherichia coli and Vibrio
XX  cholerae infections, are prepd. by solid phase peptide synthesis
XX  Claim 7; Page 11; 13pp; French.
XX  The inventors claim cholera toxin B1 subunit sequences which carry
CC  Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have
CC  shown that Arg in posns. 35, 67 and 73 play an important role in fixing
CC  these toxins to cell walls. The peptides are used in the treatment of,
CC  and vaccination against, cholera infections and animal and human
CC  infections due to E. coli (enterotoxin LT). The medicament may be
CC  administered by oral, intraperitoneal, sub-cutaneous or intravenous
CC  routes. For vaccines, pref. peptides having 13-30 AAs are injected
CC  without carriers. Unit dose when used as a medicament is 50-500mg as
XX  a vaccine 1-10mg of active cpd.
XX  Sequence 26 AA:
SQ
Query Match 100.0%; Score 39; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVPGSOH 7
DB 2 EVPGSOH 8

```

```

RESULT 9
AAP30265
ID AAP30265 standard; Protein; 26 AA.
XX
AC AAP30265;

```

```

RESULT 10
AAP50439
ID AAP50439 standard; protein; 41 AA.
XX
AC AAP50439;
XX
DT 01-JAN-1980 (first entry)
DE Network polymer which comprises a series of composite E. coli heat-
XX labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.
XX Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.
XX Synthetic.
XX W08502611-A.
XX 20-JUN-1985.
XX

```

PF 12-DEC-1984; 84WO-US02030.  
 XX 12-DEC-1983; 83US-0559469.  
 XX (SCRI-) SCRIPPS CLINIC RES.  
 PA Houghten RA;  
 XX WPI: 1985-159230/26.  
 DR  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 CC interpolyptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or  
 CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 41 AA;  
 OY  
 DB 15 EVPGSOH 21  
 OY 1 EVPGSOH 7  
 DB 15 EVPGSOH 21  
 RESULT 11  
 AAP50436  
 ID AAP50436 standard; protein; 46 AA.  
 XX  
 AC AAP50436;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Network polymer which comprises a series of composite E. coli heat-  
 DE labile toxin (LT)/heat-stable toxin (ST) polypeptide repeating units.  
 XX  
 KW Heat-labile enterotoxin; heat-stable toxin; vaccine; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO8502611-A.  
 XX  
 PD 20-JUN-1985.  
 XX  
 PF 12-DEC-1984; 84WO-US02030.  
 XX  
 PR 12-DEC-1983; 83US-0559469.  
 XX  
 PA (SCRI-) SCRIPPS CLINIC RES.  
 XX  
 PI Houghten RA;  
 XX  
 DR WPI: 1985-159230/26.  
 XX  
 XX New synthetic polypeptide(s) for part of E. coli enterotoxin(s) -  
 PT useful for vaccination of mammals against the enterotoxin(s)  
 XX  
 PS Claim 8; Page 100; 120pp; English.  
 XX  
 CC The repeating units are bonded together by intramolecular  
 CC interpolyptide cystine bonds formed between oxidized Cys residues  
 CC of the repeating units. This polypeptide may be used in the  
 CC vaccination of mammals for protection against the enterotoxins. The  
 CC composite polypeptide is made by solid phase synthesis or

CC recombinant DNA techniques. X9 is Met or Lys, X12 is Ala or Glu,  
 CC X39 is Asn or Tyr and X46 is Tyr or Asn. See also AAP50439-57.  
 XX  
 SQ Sequence 46 AA;  
 OY  
 DB 17 EVPGSOH 23  
 OY 1 EVPGSOH 7  
 DB 17 EVPGSOH 23  
 RESULT 12  
 AAP30600  
 ID AAP30600 standard; Protein; 47 AA.  
 XX  
 AC AAP30600;  
 XX  
 DT 21-APR-1992 (first entry)  
 XX  
 DE Sequence of amino acids 350-75 of the cholera toxin B1 subunit which  
 DE carries an Arg at posns. 35, 67 and 73.  
 XX  
 KW Cholera vaccine; therapy; E.coli infection; enterotoxin LT.  
 XX  
 OS Vibrio cholerae.  
 XX  
 PN EP95426-A.  
 XX  
 PD 30-NOV-1983.  
 XX  
 PF 26-MAY-1983; 83EP-0401052.  
 XX  
 PR 26-MAY-1982; 82EP-0009167.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCT.  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Milhaud G, Raulais D, Rivaille P, Siffert O, Dodin A;  
 PI Guyongruaz A, Delmas A;  
 XX  
 DR WPI: 1983-834645/49.  
 XX  
 PT Cholera toxin B, sub-unit polypeptide(s) as vaccines and  
 PT medicaments - effective against Escherichia coli and Vibrio  
 PT cholerae infections, are prepd. by solid phase peptide synthesis  
 XX  
 PS Claim 8; Page 11; 13pp; French.  
 XX  
 CC The inventors claim cholera toxin B1 subunit sequences which carry  
 CC Arg in posn. 35, 67 or 73 (AAP30264, AAP30265, AAP30600). Tests have  
 CC shown that Arg in posns. 35, 67 and 73 play an important role in fixing  
 CC these toxins to cell walls. The peptides are used in the treatment of,  
 CC and vaccination against, cholera infections and animal and human  
 CC infections due to E. coli (enterotoxin LT). The medicament may be  
 CC administered by oral, intraperitoneal, sub-cutaneous or intravenous  
 CC routes. For vaccines, pref. peptides having 15-30 AAs are injected  
 CC without carriers. Unit dose when used as a medicament is 50-500mg as  
 CC a vaccine 1-10mg of active cpd.  
 XX  
 SQ Sequence 47 AA;  
 OY  
 DB 23 EVPGSOH 29  
 OY 1 EVPGSOH 7  
 DB 23 EVPGSOH 29  
 Query Match 100.0%; Score 39; DB 4; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





```

XX 05-JAN-1999.
PD
XX
XX 22-AUG-1994; 94US-0292968.
PF
XX
XX 22-AUG-1994; 94US-0292968.
PR
XX 24-AUG-1993; 93US-0110947.
PR 31-MAY-1994; 94US-0251121.
XX
XX (UYAL-) UNITV ALBERTA.
PA
XX
XX Armstrong GD, Cockle SA, Hazes B, Klein MH, Loosmore S;
PI Oomen RP, Read RJ, Stein PE;
XX
XX WPI; 1999-105104/09.
DR
XX
XX Modifications to e.g. enzymatic activity, mitogenicity and cell
PT binding of pertussis holotoxin - by identifying interaction sites of
PT a molecule with crystalline toxin and modifying the identified site
XX
XX
XX Example 3; Fig 5; 40pp; English.
PS
XX
XX The invention relates to methods of preparing a pertussis holotoxin (PT)
CC having a modified biological activity. One method comprises identifying
CC at least 1 site in a PT that interacts with a molecule that is capable of
CC forming a complex with the holotoxin and which molecule is an effector
CC molecule which is an adenine nucleotide and which site contributes to
CC toxicity, cell binding or enzymatic activity of PT. The functional
CC interacting site(s) are identified by analysing the three dimensional
CC structure of crystalline PT, determined by X-ray crystallography. The
CC identified interacting site(s) are modified to alter toxicity, cell
CC binding or enzyme activity of the PT. The methods can be used to alter a
CC biological activity such as toxicity, enzymatic activity, mitogenicity,
CC cell binding and adjuvanticity of the PT. The three-dimensional structure
CC of PT have functional and/or structural resemblance to other bacterial
CC toxins such as diphtheria toxin (DT), Pseudomonas exotoxin A (ETA), the
CC heat-labile toxin of E. coli (LT) and verotoxin-1 (VT). The present
CC sequence represents the beta-subunit of LT toxin.
XX
XX SQ Sequence 93 AA;

Query Match 100.0%; Score 39; DB 20; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPGSOH 7
   |||||
DB 41 EYPGSOH 47

Search completed: October 24, 2002, 15:20:27
Job time : 13.2459 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 24, 2002, 15:18:53 : Search time 4.1115 Seconds  
(without alignments)  
41.388 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EYVGSQH 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/laa/5A.COMB.pep.\*

2: /cgn2\_6/ptodata/1/laa/5B.COMB.pep.\*

3: /cgn2\_6/ptodata/1/laa/6A.COMB.pep.\*

4: /cgn2\_6/ptodata/1/laa/6B.COMB.pep.\*

5: /cgn2\_6/ptodata/1/laa/PCTUS.COMB.pep.\*

6: /cgn2\_6/ptodata/1/laa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	93	2	US-08-292-968-26
2	39	100.0	93	2	US-08-467-974-26
3	39	100.0	93	2	US-08-467-974-26
4	39	100.0	93	2	US-08-467-974-26
5	39	100.0	93	2	US-08-467-974-26
6	39	100.0	93	2	US-08-467-974-26
7	39	100.0	93	2	US-08-467-974-26
8	39	100.0	93	2	US-08-467-974-26
9	39	100.0	93	2	US-08-467-974-26
10	39	100.0	93	2	US-08-467-974-26
11	39	100.0	93	2	US-08-467-974-26
12	39	100.0	93	2	US-08-467-974-26
13	39	100.0	93	2	US-08-467-974-26
14	39	100.0	93	2	US-08-467-974-26
15	39	100.0	93	2	US-08-467-974-26
16	39	100.0	93	2	US-08-467-974-26
17	39	100.0	93	2	US-08-467-974-26
18	39	100.0	93	2	US-08-467-974-26
19	39	100.0	93	2	US-08-467-974-26
20	39	100.0	93	2	US-08-467-974-26
21	39	100.0	93	2	US-08-467-974-26
22	39	100.0	93	2	US-08-467-974-26
23	39	100.0	93	2	US-08-467-974-26
24	39	100.0	93	2	US-08-467-974-26
25	39	100.0	93	2	US-08-467-974-26
26	39	100.0	93	2	US-08-467-974-26
27	39	100.0	93	2	US-08-467-974-26

28	30	76.9	329	2	US-08-781-802-8	Sequence 8, Appl
29	30	76.9	329	4	US-08-694-078-8	Sequence 8, Appl
30	30	76.9	329	4	US-09-058-260-8	Sequence 8, Appl
31	30	76.9	459	6	5194375-2	Patent No. 5194375
32	30	76.9	775	2	US-08-714-070A-1	Sequence 1, Appl
33	30	76.9	805	1	US-08-045-806-2	Sequence 2, Appl
34	30	76.9	805	1	US-08-366-051B-2	Sequence 2, Appl
35	30	76.9	855	2	US-09-027-337-2	Sequence 2, Appl
36	30	76.9	1285	2	US-08-540-406-6	Sequence 6, Appl
37	30	76.9	1285	2	US-08-656-035-6	Sequence 6, Appl
38	30	76.9	1285	4	US-08-954-668-6	Sequence 6, Appl
39	30	76.9	1285	5	PCT-US95-13233-6	Sequence 6, Appl
40	30	76.9	1286	4	US-09-268-140-3	Sequence 3, Appl
41	30	76.9	1299	4	US-08-460-900C-62	Sequence 62, Appl
42	30	76.9	1299	4	US-08-674-509B-48	Sequence 48, Appl
43	30	76.9	1299	4	US-08-954-698-48	Sequence 48, Appl
44	30	76.9	1810	5	PCT-US95-11684-4	Sequence 4, Appl
45	29	74.4	346	2	US-08-602-359A-34	Sequence 34, Appl

## ALIGNMENTS

RESULT 1  
US-08-292-968-26  
Sequence 26, Application US/08292968  
Patent No. 5656122  
GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & Mcburney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,968  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-388  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-292-968-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
|||||  
DB 41 EYVGSQH 47

RESULT 2

US-08-467-974-26  
Sequence 26, Application US/08467974  
Patent No. 5965385

GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COHEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,974  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/467,536  
FILING DATE: 06-JUN-1995  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-454 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-974-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7

DB 41 EYVGSQH 47  
|||||

RESULT 3

US-08-467-536-26  
Sequence 26, Application US/08467536  
Patent No. 5977304

GENERAL INFORMATION:  
APPLICANT: READ, Randy J.  
APPLICANT: STEIN, Penelope E.  
APPLICANT: COCKLE, Stephen A.  
APPLICANT: COHEN, Raymond P.  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: KLEIN, Michel H.  
APPLICANT: ARMSTRONG, Glen D.  
APPLICANT: HAZES, Bart  
TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,536  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/292,968  
FILING DATE: 22-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/251,121  
FILING DATE: 31-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,947  
FILING DATE: 24-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-455 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 93 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-467-536-26

Query Match 100.0%; Score 39; DB 2; Length 93;  
Best Local Similarity 100.0%; Pred. No. 0.67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
|||||  
DB 41 EYVGSQH 47

RESULT 4  
US-08-467-976-26  
Sequence 26, Application US/08467976  
Patent No. 6018022  
GENERAL INFORMATION:

```

? APPLICANT: READ, Randy J.
? APPLICANT: STEIN, Penelope E.
? APPLICANT: COCKLE, Stephen A.
? APPLICANT: OOMEN, Raymond P.
? APPLICANT: LOOSMORE, Sheena
? APPLICANT: KLEIN, Michel H.
? APPLICANT: ARMSTRONG, Glen D.
? APPLICANT: HAZES, Bart
? TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
? NUMBER OF SEQUENCES: 46
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sim & Mcburney
? STREET: Suite 701, 330 University Avenue
? City: Toronto
? STATE: Ontario
? COUNTRY: Canada
? ZIP: M5G 1R7
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/467, 976
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/292, 968
? FILING DATE: 22-AUG-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/251, 121
? FILING DATE: 31-MAY-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/110, 947
? FILING DATE: 24-AUG-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: STEWART, Michael I.
? REGISTRATION NUMBER: 24, 973
? REFERENCE/DOCKET NUMBER: 1038-453 MIS:v9
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 26:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 93 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-467-976-26

Query Match      100.0%; Score 39; DB 3; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  EVPGSOH 7
DB      41  EVPGSOH 47

RESULT 5
US-09-082-514-26
? Sequence 26, Application US/09082514
? Patent No. 6168928
? GENERAL INFORMATION:
? APPLICANT: READ, Randy J.
? APPLICANT: STEIN, Penelope E.
? APPLICANT: COCKLE, Stephen A.
? APPLICANT: OOMEN, Raymond P.
? APPLICANT: KLEIN, Michel H.
? APPLICANT: ARMSTRONG, Glen D.
? APPLICANT: HAZES, Bart
? TITLE OF INVENTION: MODIFICATION OF PERTUSSIS TOXIN
? NUMBER OF SEQUENCES: 46

```

```

? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sim & Mcburney
? STREET: 6th Floor, 330 University Avenue
? City: Toronto
? STATE: Ontario
? COUNTRY: Canada
? ZIP: M5G 1R7
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/082, 514
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/292, 968
? FILING DATE: 22-AUG-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: STEWART, Michael I.
? REGISTRATION NUMBER: 24, 973
? REFERENCE/DOCKET NUMBER: 1038-810
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 26:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 93 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-09-082-514-26

Query Match      100.0%; Score 39; DB 4; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  EVPGSOH 7
DB      41  EVPGSOH 47

RESULT 6
US-08-952-337-5
? Sequence 5, Application US/08952337
? Patent No. 6019973
? GENERAL INFORMATION:
? APPLICANT: Holmgren, Jan
? APPLICANT: Lebens, Michael R.
? TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
? TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
? FILE REFERENCE: 3846/OD758
? CURRENT APPLICATION NUMBER: US/08/952, 337
? CURRENT FILING DATE: 1998-01-05
? EARLIER APPLICATION NUMBER: PCT/SE96/00570
? EARLIER FILING DATE: 1996-05-02
? EARLIER APPLICATION NUMBER: SE 9501682-0
? EARLIER FILING DATE: 1995-05-05
? NUMBER OF SEQ ID NOS: 6
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 5
? LENGTH: 102
? TYPE: PRT
? ORGANISM: Vibrio cholerae
?
US-08-952-337-5

Query Match      100.0%; Score 39; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  EVPGSOH 7
DB      41  EVPGSOH 47

```

DB 50 EYVGSOH 56

RESULT 7  
US-08-952-337-6  
; Sequence 6, Application US/08952337  
; Patent No. 6019973  
; GENERAL INFORMATION:  
; APPLICANT: Holmgren, Jan  
; APPLICANT: Lebens, Michael R.  
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE  
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS  
; FILE REFERENCE: 3846/OD758  
; CURRENT APPLICATION NUMBER: US/08/952,337  
; CURRENT FILING DATE: 1998-01-05  
; EARLIER APPLICATION NUMBER: PCT/SE96/00570  
; EARLIER FILING DATE: 1996-05-02  
; EARLIER APPLICATION NUMBER: SE 9501682-0  
; EARLIER FILING DATE: 1995-05-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-08-952-337-6

Query Match 100.0%; Score 39; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 0.73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSOH 7  
|||||||  
DB 50 EYVGSOH 56

RESULT 8  
US-08-472-171-2  
; Sequence 2, Application US/084722171  
; Patent No. 5932714  
; GENERAL INFORMATION:  
; APPLICANT: Loomore, Sheena M.  
; APPLICANT: Yacoub, Reza K.  
; APPLICANT: Zealey, Gavin R.  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Expression Of Gene Products From  
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 330 University Avenue, Suite 701  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,171  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/393,334  
; FILING DATE: 23-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I.  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 416-595-1155

; TELEFAX: 416-595-1163  
; TELEX: 065-24567 Simdas  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-472-171-2

Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSOH 7  
|||||||  
DB 51 EYVGSOH 57

RESULT 9  
US-08-894-526-2  
; Sequence 2, Application US/08894526  
; Patent No. 5942418  
; GENERAL INFORMATION:  
; APPLICANT: Loomore, Sheena M  
; APPLICANT: Yacoub, Reza K  
; APPLICANT: Zealey, Gavin R  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: EXPRESSION OF GENE PRODUCTS FROM  
; TITLE OF INVENTION: GENETICALLY MANIPULATED STRAINS OF BORDETELLA  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,526  
; FILING DATE: 01-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-724 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-894-526-2

Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSOH 7  
|||||||  
DB 51 EYVGSOH 57

RESULT 10  
US-09-013-047-2

```

; Sequence 2, Application US/09013047
; Patent No. 5998168
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michael H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,047
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,171
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,334
; FILING DATE: 23-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-507 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-047-2

Query Match          100.0%; Score 39; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. NO. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 EVPGSQH 7
        |||||
Db      51 EVPGSQH 57

```

```

RESULT 11
; Sequence 2, Application US/09374597
; Patent No. 614082
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yacoub, Reza K.
; APPLICANT: Zealey, Gavin R.
; APPLICANT: Klein, Michael H.
; TITLE OF INVENTION: Expression Of Gene Products From
; TITLE OF INVENTION: Genetically Manipulated Strains Of Bordetella
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; City: Toronto

```

```

; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/393,334
; FILING DATE: FEBRUARY 23, 1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-595-1155
; TELEFAX: 416-595-1163
; TELEX: 065-24567 Simbas
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-374-597-2

Query Match          100.0%; Score 39; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. NO. 0.74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 EVPGSQH 7
        |||||
Db      51 EVPGSQH 57

```

```

RESULT 12
; Sequence 21, Application US/09191852
; Patent No. 6194560
; GENERAL INFORMATION:
; APPLICANT: Charles J. Arnizen, Hugh S. Mason, and Tariq A. Haq
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 1301 McKinney, Suite 5100
; City: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,852
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13376
; FILING DATE: 24-OCT-1995
; APPLICATION NUMBER: 08/817,906
; FILING DATE: 04-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, David L.
; REGISTRATION NUMBER: 40,612

```

```

; REFERENCE/DOCKET NUMBER: P01590US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5151
; TELEFAX: 713-651-5246
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-191-852-21

Query Match
Best Local Similarity 100.0%; Score 39; DB 4; Length 103;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   1111111
Db 51 EYVGSQH 57

RESULT 13
PCT-US95-13376-21
; Sequence 21, Application PC/TUS9513376
; GENERAL INFORMATION:
; APPLICANT: The Texas A&M University System
; APPLICANT: 310 Wisenbaker
; APPLICANT: College Station, Texas 77843-3369
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1177 West Loop South, 10th Floor
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77027-9095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13376
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/328,716
; FILING DATE: 24-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jones, John W.
; REGISTRATION NUMBER: 31,380
; REFERENCE/DOCKET NUMBER: 36170/3P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0909
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
PCT-US95-13376-21

Query Match
Best Local Similarity 100.0%; Score 39; DB 5; Length 103;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   1111111
Db 51 EYVGSQH 57

RESULT 14
US-08-952-337-1
```

```

; Sequence 1, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-08-952-337-1

Query Match
Best Local Similarity 100.0%; Score 39; DB 3; Length 123;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   1111111
Db 71 EYVGSQH 77

RESULT 15
US-08-952-337-2
; Sequence 2, Application US/08952337
; Patent No. 6019973
; GENERAL INFORMATION:
; APPLICANT: Holmgren, Jan
; APPLICANT: Lebens, Michael R.
; TITLE OF INVENTION: HYBRID MOLECULES BETWEEN HEAT-LABILE
; TITLE OF INVENTION: ENTEROTOXIN AND CHOLERA TOXIN B SUBUNITS
; FILE REFERENCE: 3846/0D758
; CURRENT APPLICATION NUMBER: US/08/952,337
; CURRENT FILING DATE: 1998-01-05
; EARLIER APPLICATION NUMBER: PCT/SE96/00570
; EARLIER FILING DATE: 1996-05-02
; EARLIER APPLICATION NUMBER: SE 9501682-0
; EARLIER FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-952-337-2

Query Match
Best Local Similarity 100.0%; Score 39; DB 3; Length 123;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   1111111
Db 71 EYVGSQH 77

Search completed: October 24, 2002, 15:23:57
Job time : 6.1315 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:21:08 : Search time 15.8361 Seconds  
(without alignments)  
132.044 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BLOSUM62

Searched: Gapop 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 1105779

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	103	5	US-09-791-537-19387
2	39	100.0	103	5	US-09-791-537-38611
3	39	100.0	103	5	US-09-791-537-38639
4	39	100.0	103	5	US-09-791-537-68581
5	39	100.0	103	5	US-09-791-537-74385
6	39	100.0	103	6	US-10-110-364-8
7	39	100.0	103	6	US-10-110-364-10
8	39	100.0	103	6	US-10-110-364-13
9	39	100.0	103	6	US-10-110-364-22
10	39	100.0	104	5	US-09-791-537-87980
11	39	100.0	104	5	US-09-791-537-99772
12	39	100.0	105	6	US-10-110-364-12
13	39	100.0	113	5	US-09-791-537-73608
14	39	100.0	123	6	US-10-110-364-17
15	39	100.0	123	6	US-10-110-364-20
16	39	100.0	124	5	US-09-791-537-28360
17	39	100.0	124	5	US-09-791-537-29489
18	39	100.0	124	5	US-09-791-537-33623
19	39	100.0	124	5	US-09-791-537-40552
20	39	100.0	124	5	US-09-791-537-60743
21	39	100.0	124	5	US-09-791-537-78640
22	39	100.0	124	5	US-09-791-537-92185
23	39	100.0	124	5	US-09-791-537-94644
24	39	100.0	124	5	US-09-791-537-103241
25	39	100.0	124	5	US-09-791-537-103660
26	39	100.0	124	6	US-10-110-364-2

27	39	100.0	124	6	US-10-110-364-6	Sequence 6, Appl1
28	39	100.0	124	6	US-10-110-364-11	Sequence 11, Appl
29	39	100.0	124	6	US-10-110-364-16	Sequence 16, Appl
30	39	100.0	124	6	US-10-110-364-18	Sequence 18, Appl
31	39	100.0	124	6	US-10-110-364-19	Sequence 19, Appl
32	39	100.0	124	6	US-10-110-364-21	Sequence 21, Appl
33	39	100.0	124	6	US-10-110-364-23	Sequence 23, Appl
34	39	100.0	125	6	US-10-110-364-15	Sequence 15, Appl
35	39	100.0	129	5	US-09-791-537-131854	Sequence 131854,
36	39	100.0	131	5	US-09-791-537-130348	Sequence 130348,
37	39	100.0	138	6	US-10-141-627-2	Sequence 2, Appl1
38	36	92.3	103	5	US-09-791-537-42610	Sequence 42610, A
39	36	92.3	103	6	US-10-110-364-7	Sequence 7, Appl1
40	35	89.7	565	5	US-09-791-537-47818	Sequence 47818, A
41	33	84.6	103	5	US-09-791-537-129309	Sequence 129309,
42	33	84.6	103	6	US-10-110-364-5	Sequence 5, Appl1
43	33	84.6	124	5	US-09-791-537-123948	Sequence 123948,
44	33	84.6	124	6	US-10-110-364-4	Sequence 4, Appl1
45	33	84.6	124	6	US-10-110-364-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-791-537-19387  
Sequence 19387, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19387  
LENGTH: 103  
TYPE: PRT  
ORGANISM: pdb 1CPIID  
US-09-791-537-19387

Query Match 100.0%; Score 39; DB 5; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQH 7  
|||||||  
Db 51 EVPGSQH 57

RESULT 2  
US-09-791-537-38611  
Sequence 38611, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 38611  
LENGTH: 103  
TYPE: PRT  
ORGANISM: pdb 1CHPD  
US-09-791-537-38611

```
Query Match          100.0%; Score 39; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   |||||
Db 51 EYVGSQH 57

RESULT 3
US-09-791-537-38639
; Sequence 38639, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38639
; LENGTH: 103
; TYPE: PRT
; ORGANISM: pdb 1CHOD
US-09-791-537-38639

Query Match          100.0%; Score 39; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   |||||
Db 51 EYVGSQH 57

RESULT 4
US-09-791-537-68591
; Sequence 68591, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 68591
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-09-791-537-68591

Query Match          100.0%; Score 39; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   |||||
Db 51 EYVGSQH 57

RESULT 5
US-09-791-537-74385
; Sequence 74385, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek

Query Match          100.0%; Score 39; DB 5; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   |||||
Db 51 EYVGSQH 57

RESULT 6
US-10-110-364-8
; Sequence 8, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/158,561
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Vibrio cholera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)-(103)
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 998409
US-10-110-364-8

Query Match          100.0%; Score 39; DB 6; Length 103;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVGSQH 7
   |||||
Db 51 EYVGSQH 57

RESULT 7
US-10-110-364-10
; Sequence 10, Application US/10110364
; GENERAL INFORMATION:
; APPLICANT: Handley, Harold H.
; APPLICANT: Haaparanta, Tapio
; APPLICANT: Ewalt, Karla L.
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED
; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS
; FILE REFERENCE: ACTBIO.004A
; CURRENT APPLICATION NUMBER: US/10/110,364
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: PCT/US00/27607
```

;; PRIOR FILING DATE: 2000-10-05  
;; PRIOR APPLICATION NUMBER: 60/158,561  
;; PRIOR FILING DATE: 1999-10-08  
;; NUMBER OF SEQ ID NOS: 31  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 10  
;; LENGTH: 103  
;; TYPE: PRF  
;; ORGANISM: Vibrio cholera  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (1)...(103)  
;; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 1421511  
US-10-110-364-10

Query Match 100.0%; Score 39; DB 6; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 51 EVPGSQH 57

RESULT 8  
US-10-110-364-13  
;; Sequence 13, Application US/10110364  
;; GENERAL INFORMATION:  
;; APPLICANT: Handley, Harold H.  
;; APPLICANT: Haaparanta, Tapio  
;; APPLICANT: Ewalt, Karla L.  
;; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
;; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS  
;; FILE REFERENCE: ACTBIO.004A  
;; CURRENT APPLICATION NUMBER: US/10/110,364  
;; CURRENT FILING DATE: 2002-04-05  
;; PRIOR APPLICATION NUMBER: PCT/US00/27607  
;; PRIOR FILING DATE: 2000-10-05  
;; PRIOR APPLICATION NUMBER: 60/158,561  
;; PRIOR FILING DATE: 1999-10-08  
;; NUMBER OF SEQ ID NOS: 31  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 13  
;; LENGTH: 103  
;; TYPE: PRF  
;; ORGANISM: Vibrio cholera  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (1)...(103)  
;; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 14215235  
US-10-110-364-13

Query Match 100.0%; Score 39; DB 6; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 51 EVPGSQH 57

RESULT 9  
US-10-110-364-22  
;; Sequence 22, Application US/10110364  
;; GENERAL INFORMATION:  
;; APPLICANT: Handley, Harold H.  
;; APPLICANT: Haaparanta, Tapio  
;; APPLICANT: Ewalt, Karla L.  
;; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
;; TITLE OF INVENTION: CHEMICAL CONJUGATION CHARACTERISTICS  
;; FILE REFERENCE: ACTBIO.004A  
;; CURRENT APPLICATION NUMBER: US/10/110,364

;; CURRENT FILING DATE: 2002-04-05  
;; PRIOR APPLICATION NUMBER: PCT/US00/27607  
;; PRIOR FILING DATE: 2000-10-05  
;; PRIOR APPLICATION NUMBER: 60/158,561  
;; PRIOR FILING DATE: 1999-10-08  
;; NUMBER OF SEQ ID NOS: 31  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 22  
;; LENGTH: 103  
;; TYPE: PRF  
;; ORGANISM: Escherichia coli  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (1)...(103)  
;; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 494265.  
US-10-110-364-22

Query Match 100.0%; Score 39; DB 6; Length 103;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 51 EVPGSQH 57

RESULT 10  
US-09-791-537-87980  
;; Sequence 87980, Application US/09791537  
;; GENERAL INFORMATION:  
;; APPLICANT: Bionomix, Inc.  
;; APPLICANT: Danzer, Joseph  
;; APPLICANT: Debe, Derek  
;; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
;; TITLE OF INVENTION: METHODS OF USE THEREOF  
;; FILE REFERENCE: 261/210  
;; CURRENT APPLICATION NUMBER: US/09/791,537  
;; CURRENT FILING DATE: 2001-02-22  
;; NUMBER OF SEQ ID NOS: 153055  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 87980  
;; LENGTH: 104  
;; TYPE: PRF  
;; ORGANISM: pdb 2CHBD  
US-09-791-537-87980

Query Match 100.0%; Score 39; DB 5; Length 104;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7  
Db 52 EVPGSQH 58

RESULT 11  
US-09-791-537-99772  
;; Sequence 99772, Application US/09791537  
;; GENERAL INFORMATION:  
;; APPLICANT: Bionomix, Inc.  
;; APPLICANT: Danzer, Joseph  
;; APPLICANT: Debe, Derek  
;; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
;; TITLE OF INVENTION: METHODS OF USE THEREOF  
;; FILE REFERENCE: 261/210  
;; CURRENT APPLICATION NUMBER: US/09/791,537  
;; CURRENT FILING DATE: 2001-02-22  
;; NUMBER OF SEQ ID NOS: 153055  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 99772  
;; LENGTH: 104  
;; TYPE: PRF  
;; ORGANISM: pdb 3CHBD

US-09-791-537-99772

Query Match 100.0%; Score 39; DB 5; Length 104;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7  
|||||||  
DB 52 EVPGSOH 58

RESULT 12

US-10-110-364-12  
; Sequence 12, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 105  
; TYPE: PRT  
; ORGANISM: Vibrio cholera  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(105)  
; OTHER INFORMATION: CTB variant from NCBI gene bank GI: 2781121 (ogawa  
; US-10-110-364-12

Query Match 100.0%; Score 39; DB 6; Length 105;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7  
|||||||  
DB 52 EVPGSOH 58

RESULT 13

US-09-791-537-73608  
; Sequence 73608, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danczer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; FILE REFERENCE: 261/210  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 73608  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: dbb 1LTRD  
; US-09-791-537-73608

Query Match 100.0%; Score 39; DB 5; Length 113;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7

DB 51 EVPGSOH 57

RESULT 14

US-10-110-364-17  
; Sequence 17, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(123)  
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 1395122.  
; US-10-110-364-17

Query Match 100.0%; Score 39; DB 6; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7  
|||||||  
DB 72 EVPGSOH 78

RESULT 15

US-10-110-364-20  
; Sequence 20, Application US/10110364  
; GENERAL INFORMATION:  
; APPLICANT: Handley, Harold H.  
; APPLICANT: Haaparanta, Tapio  
; APPLICANT: Ewalt, Karla L.  
; TITLE OF INVENTION: ABS TOXIN B SUBUNIT MUTANTS WITH ALTERED  
; FILE REFERENCE: ACTBIO.004A  
; CURRENT FILING DATE: 2002-04-05  
; PRIOR APPLICATION NUMBER: PCT/US00/27607  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 60/158,561  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(123)  
; OTHER INFORMATION: LTB variant from NCBI gene bank GI: 223254.  
; US-10-110-364-20

Query Match 100.0%; Score 39; DB 6; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Fri Oct 25 17:50:53 2002

us-09-786-648-2.rapn

Page 5

OY 1 EYPGSGH 7  
|||||  
Db 71 EYPGSGH 77

Search completed: October 24, 2002, 15:33:19  
Job time : 16.8361 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:17:58 : Search time 5.39344 Seconds  
(without alignments)  
124.712 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPSQH 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-71:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	124	1 XYVCB	Cholera enterotoxin
2	39	100.0	124	1 QLECB	heat-labile entero
3	33	84.6	272	2 T14755	hypothetical prote
4	33	84.6	367	2 T01751	gibberellin 20-oxi
5	32	82.1	137	2 G96017	conserved hypothet
6	32	82.1	455	1 A69753	glucanate dehydrat
7	32	82.1	472	2 AG0432	glutamate synthase
8	32	82.1	534	2 S44886	ZK112.1 protein -
9	31	79.5	137	2 B27586	hypothetical prote
10	31	79.5	242	2 T34767	hypothetical prote
11	31	79.5	260	2 C96827	protein F20B1.2 (
12	31	79.5	273	2 B48820	homeobox protein (
13	31	79.5	274	2 G84353	hypothetical prote
14	31	79.5	353	2 H75446	(S)-2-hydroxy-acid
15	31	79.5	374	2 DB1715	conserved hypothet
16	31	79.5	414	2 A37133	apolipoprotein A-I
17	31	79.5	414	2 I48975	apolipoprotein A-I
18	31	79.5	432	2 T46725	chorismate synthas
19	31	79.5	432	2 T51020	chorismate synthas
20	31	79.5	459	2 D34791	interleukin-7 rece
21	31	79.5	489	2 T26069	hypothetical prote
22	31	79.5	703	2 AG0242	probable membrane
23	31	79.5	708	2 AG2315	hypothetical prote
24	31	79.5	755	2 D95842	probable beta-gala
25	31	79.5	1028	2 A96719	hypothetical prote
26	31	79.5	1228	2 C98219	proline dehydrogen
27	31	79.5	1228	2 AG3067	proline dehydrogen
28	31	79.5	1258	2 T29041	hypothetical prote
29	31	79.5	1259	2 A43425	Bravo/Nr-CAM cell

30	30	76.9	148	2 B72782	hypothetical prote
31	30	76.9	153	2 T31701	hypothetical prote
32	30	76.9	182	2 D83638	conserved hypothet
33	30	76.9	262	2 C34791	interleukin-7 rece
34	30	76.9	270	2 D83072	conserved hypothet
35	30	76.9	291	2 A12241	hypothetical prote
36	30	76.9	298	2 B34791	interleukin-7 rece
37	30	76.9	304	2 F84169	hypothetical prote
38	30	76.9	336	2 A72247	DNA-directed RNA p
39	30	76.9	373	2 T47115	probable 4-carboxy
40	30	76.9	375	2 T35015	probable 3-oxoadip
41	30	76.9	376	1 S17246	chorismate synthas
42	30	76.9	412	2 T47142	hypothetical prote
43	30	76.9	431	2 A89761	hypothetical prote
44	30	76.9	436	2 B70321	flavocytochrome C
45	30	76.9	440	2 T44138	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

XYVCB

Cholera enterotoxin, B chain precursor VC1456 [validated] - Vibrio cholerae (strain N  
N:Alternate names: enterotoxin beta chain

C:Species: Vibrio cholerae

C:Date: 24-Apr-1984 #sequence-revision 01-Sep-2000 #text-change 02-Feb-2001

C:Accession: S14624; S39238; S39241; B82196; J01078; S17666; P01010; A05130; A01819;

R:Dams, E.; de Wolf, M.; Dierick, W.

submitted to the EMBL Data Library, March 1991

A:Description: Correction of the cholera toxin nucleotide sequence of the Vibrio chol

A:Reference number: S14623

A:Accession: S14624

A:Molecule type: DNA

A:Residues: 1-124 <DAM>

A:Cross-references: EMBL:X58786; NID:948420; PIDN:CAA41593.1; PID:948422

A:Experimental source: strain 2125

R:Lebens, M.; Holmgren, J.

submitted to the EMBL Data Library, November 1993

A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera

A:Reference number: S39238

A:Accession: S39238

A:Molecule type: DNA

A:Residues: 1-124 <LEB>

A:Cross-references: EMBL:X76390; NID:9433856; PIDN:CAA53973.1; PID:9433857

A:Accession: S39241

A:Molecule type: DNA

A:Residues: 1-124 <LEM>

A:Cross-references: EMBL:X76391; NID:9433859; PIDN:CAA53976.1; PID:9433861

R:Heidelberg, J.F.; Eisen, J.A.; Neilson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

charlson, D.; Emolaeva, M.D.; Yamathayan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: H82196

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-124 <HEI>

A:Cross-references: GB:AE004224; GB:AE003852; NID:99655952; PIDN:AAF94613.1; GSPDB:GN

A:Experimental source: serogroup O1, strain N16961; biotype El Tor

R:Shi, C.H.; Cao, C.; Zhang, J.S.; Ma, Q.J.

Chinese Biochem J. 9, 395-399, 1993

A:Title: Nucleotide sequence analysis of the gene encoding the classical biotype V.ch

A:Reference number: J01078

A:Accession: J01078

A:Molecule type: DNA

A:Residues: 1-20, 'O', 22-31, 'O', 33-38, 'H', 40-49, 'G', 51-67, 'T', 69-124 <SHI>

A:Experimental source: classical biotype strain 569B

R:Dams, E.; de Wolf, M.; Dierick, W.

Biochem. Biophys. Acta 1090, 139-141, 1991

A:Title: Nucleotide sequence analysis of the CT operon of the Vibrio cholerae classic

A:Reference number: S17665; MUID:91355224

A:Accession: S17666  
 A:Molecule type: DNA  
 A:Residues: 1-38, 'H', '40-67', 'T', '69-124 <DA2>  
 A:Cross-references: EMBL:X58785; NID:g48888; PIDD:CAA41591.1; PID:g48890  
 R:Ma, Q.T.; Liu, C.X.; Xiong, L.S.; Yu, X.Q.  
 Sci. Sin. B Chem. Biol. Agric. Med. Earth Sci. 34, 274-280, 1991  
 A:Title: B subunit of cholera toxin produced in *Escherichia coli*.  
 A:Reference number: PC1010  
 A:Accession: PC1010  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', '40-41 <MAO>  
 R:Melakatos, J.J.; Swartz, D.J.; Pearson, G.D.N.; Harford, N.; Groyne, F.; de Wilde, M.  
 Nature 306, 551-557, 1983  
 A:Reference number: A93320; MWID:84068199  
 A:Accession: A05130  
 A:Molecule type: DNA  
 A:Residues: 1-32, 'S', '34-74', 'S', '76-124 <MEK>  
 A:Cross-references: GB:X00171; NID:g48347; PIDD:CAA24996.1; PID:g758351  
 R:Kurosky, A.; Markel, D.E.; Peterson, J.W.  
 J. Biol. Chem. 252, 7257-7264, 1977  
 A:Title: Covalent structure of the beta chain of cholera enterotoxin.  
 A:Reference number: A01819; MWID:78005537  
 A:Accession: A01819  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', '40-42', 'N', '44-67', 'T', '69-90', 'N', '92-124 <KUR>  
 R:Mal, C.Y.  
 J. Biol. Chem. 252, 7249-7256, 1977  
 A:Title: Determination of the primary structure of cholera toxin B subunit.  
 A:Reference number: A38033; MWID:78005536  
 A:Accession: A38033  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', '40-42', 'N', '44-67', 'T', '69', 'E', '71-90', 'N', '92-124 <LAI>  
 A:Note: the difference at residue 70 may be due to deamidation during preparation  
 R:Nakashima, Y.; Napietkowski, P.; Schaefer, D.E.; Konigsberg, W.H.  
 FEBS Lett. 68, 275-278, 1976  
 A:Title: Primary structure of the B subunit of cholera enterotoxin.  
 A:Reference number: A38034; MWID:77026365  
 A:Accession: A38034  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', '40-67', 'T', '69', 'E', '71', 'QS', '74-75', 'VE', '78-86', 'Q', '88-99', 'Q', '101-103', 'Q'  
 R:Takao, T.; Matnabe, H.; Shimonishi, Y.  
 Eur. J. Biochem. 146, 503-508, 1985  
 A:Title: Facile identification of protein sequences by mass spectrometry.  
 A:Reference number: A21910; MWID:85126976  
 A:Accession: A21910  
 A:Molecule type: protein  
 A:Residues: 22-38, 'H', '40-42', 'N', '44-67', 'T', '69-90', 'N', '92-124 <TAK>  
 A:Experimental source: biotype Inaba 569B  
 C:Note: Asn-65 was partially deaminated to Asp  
 C:Comment: The authors translated the codon TCA for residue 33 as Tyr.  
 C:Genetics:  
 A:Gene: VC1456  
 A:Map position: 1  
 C:Complex: the cholera enterotoxin molecule contains three kinds of chains; an alpha and  
 ciate noncovalently with the subunit B, an aggregate of five beta chains  
 C:Function:  
 A:Description: Involved in binding of the toxin to cell membranes  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin; toxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: cholera enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status experimental

Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPSOH 7  
 |||||  
 DB 72 EYPSOH 78

RESULT 2

OLECB  
 heat-labile enterotoxin chain B precursor - *Escherichia coli*  
 C:Species: *Escherichia coli*  
 C:Date: 29-Jun-1981 #sequence, revision 29-Jun-1981 #text, change 18-Jun-1999  
 C:Accession: A01820; B26946; I41194; I41287; I67644; A61475  
 R:Dallas, W.S.; Falkow, S.  
 Nature 288, 499-501, 1980  
 A:Title: Amino acid sequence homology between cholera toxin and *Escherichia coli* heat  
 A:Reference number: A01820; MWID:81074965  
 A:Accession: A01820  
 A:Molecule type: mRNA  
 A:Residues: 1-124 <DAL>  
 R:Yamamoto, T.; Gotohori, T.; Yokota, T.  
 J. Bacteriol. 169, 1352-1357, 1987  
 A:Title: Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia*  
 A:Reference number: A26946; MWID:87137303  
 A:Accession: B26946  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'E', '29-63', 'K', '65-124 <YAM>  
 A:Cross-references: EMBL:M5363; NID:g148335; PIDD:AAA24792.1; PID:g148336  
 R:Leong, J.; Vinal, A.C.; Dallas, W.S.  
 Infect. Immun. 48, 73-77, 1985  
 A:Title: Nucleotide sequence comparison between heat-labile toxin B subunit cistrons  
 A:Reference number: I41194; MWID:85156481  
 A:Accession: I41194  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-5, 'F', '7-17', 'C', '19-24', 'S', '26-27', 'E', '29-33', 'H', '35-63', 'K', '65-66', 'A', '68-122  
 A:Cross-references: GB:M17874; NID:g145830; PIDD:AAA98064.1; PID:g145831  
 R:Experimental source: Plasmid ENT-R PCG86  
 R:Ibrahim, I.; Gentz, R.  
 J. Biol. Chem. 262, 10189-10194, 1987  
 A:Title: A functional interaction between the signal peptide and the translation appa  
 ticulum.  
 A:Reference number: I41287; MWID:87280041  
 A:Accession: I41287  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-22 <R22>  
 A:Cross-references: GB:M17101; NID:g146375; PIDD:AAA23973.1; PID:g146376  
 R:Hou, T.; Tsuji, T.; Koto, M.; Imamura, S.; Miyama, A.  
 FEBS Microbiol. Lett. 108, 157-161, 1993  
 A:Title: Amino acid sequence of heat-labile enterotoxin from chicken enterotoxigenic  
 A:Reference number: I53542; MWID:93252225  
 A:Accession: I53542  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-17, 'C', '19', 'Y', '21-24', 'S', '26-27', 'E', '29-63', 'K', '65-66', 'A', '68-122', 'E', '124 <R  
 A:Cross-references: GB:S60731; NID:g408994; PIDD:AA60441.1; PID:g408996  
 R:Tsuji, T.; Lida, T.; Honda, T.; Miyataki, T.; Nagahama, M.; Sakurai, J.; Wada, K.;  
 Microb. Pathog. 2, 381-390, 1987  
 A:Title: A unique amino acid sequence of the B subunit of a heat-labile enterotoxin 1  
 A:Reference number: A61475; MWID:89180953  
 A:Accession: A61475  
 A:Molecule type: protein  
 A:Residues: 22-24, 'S', '26-27', 'E', '29-63', 'K', '65-66', 'A', '68-95', 'A', '97-122', 'E', '124 <TSU>  
 A:Experimental source: strain 240-3  
 C:Complex: the heat-labile enterotoxin molecule contains one A chain and five or six  
 C:Function:  
 A:Description: the biological activity of the toxin is produced by the A chain, which  
 C:Superfamily: cholera enterotoxin beta chain  
 C:Keywords: enterotoxin  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-124/Product: heat-labile enterotoxin chain B #status predicted <MAT>  
 F:30-107/Disulfide bonds: #status predicted

Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYPSOH 7  
 |||||  
 DB 72 EYPSOH 78



## RESULT 3

hypothetical protein DKFZp564A0122.1 - human

C:Species: Homo sapiens (man)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14755

R:Wambolt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A/Accession: T14755

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-272 <N>

A/Cross-references: EMBL:AL110209

A/Experimental source: fetal brain; clone DKFZp564A0122

C/Genetics:

A>Note: DKFZp564A0122.1

Query Match

Best Local Similarity 84.6%; Score 33; DB 2; Length 272;

Pred. NO. 22;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 246 EDPGSEH 252

## RESULT 4

T01751

giberellin 20-oxidase - common tobacco

N/Alternate names: Ntcl6 protein

C/Species: Nicotiana tabacum (common tobacco)

C>Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 20-Jun-2000

C/Accession: T01751

R:Tanaka-Ideguchi, M.; Itoh, H.; Oyama, N.; Koshioke, M.; Matsuoka, M.

submitted to the EMBL Data Library, July 1998

A>Description: Overexpression of a tobacco homeobox gene, NTH15, decreases the expression

A/Reference number: Z14418

A/Accession: T01751

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-367 <N>

A/Cross-references: EMBL:AB016084

C/Genetics:

A/Gene: Ntcl6

C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match

Best Local Similarity 85.7%; Score 33; DB 2; Length 367;

Pred. NO. 30;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 149 EDPGSEH 155

## RESULT 5

G96017

conserved hypothetical exported protein SMB20700 [imported] - Sinorhizobium meliloti (st

C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C/Accession: G96017

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A>Title: The complete sequence of the 1.683-kb PSymb megaplasmid from the N2-fixing endo

A/Reference number: A95842; MUID:21396508; PMID:11481431

A/Accession: G96017

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-137 <N>

A/Cross-references: GB:AL591965; PIDN:GAC49807.1; PID:G15141295; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubli

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Drano, S.; Federspiel, N.A.; Fisher, R.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Jelan

heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,

A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A/Reference number: A96039; MUID:21368234; PMID:11474104

A/Contents: annotation

C/Genetics:

A/Gene: SMB20700

A/Genome: plasmid

Query Match

Best Local Similarity 82.1%; Score 32; DB 2; Length 137;

Pred. NO. 17;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 98 EDPGSEH 104

## RESULT 6

A69753

glucarate dehydratase (EC 4.2.1.40) - Bacillus subtilis

C/Species: Bacillus subtilis

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C/Accession: A69753

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Poulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadleir, Y.; Sato, T.; Scari

A:Authors: Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se

akenchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchida

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A/Reference number: A69580; MUID:96044033

A/Accession: A69753

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-455 <N>

A/Cross-references: GB:299105; GB:AL009126; NID:G2632457; PIDN:CAB12043.1; PID:G26325

C:Experimental source: strain 168

C/Genetics:

A/Gene: ycbF

C:Superfamily: glucarate dehydratase

C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match

Best Local Similarity 82.1%; Score 32; DB 1; Length 455;

Pred. NO. 62;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7

Db 64 EDPGSEH 70

## RESULT 7

AG0432

glutamate synthase (NADPH) (EC 1.4.1.13) small chain [imported] - Yersinia pestis (st

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001

C/Accession: AG0432

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G



A:Title: Two distal-less related homeobox-containing genes expressed in regeneration bla  
 A:Reference number: A48820; MUID:93050784  
 A:Accession: B48820  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-273 <BEA>  
 A:Cross-references: GB:X63531; GB:S47223; NID:q432377; PIDN:CAAA5094.1; PID:q432378  
 A:Note: sequence extracted from NCBI backbone (NCBIN:117052, NCBI:P:117053)  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:126-182/Domain: homeobox homology <HOX>

Query Match  
 Best Local Similarity 79.5%; Score 31; DB 2; Length 273;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 |||||  
 Db 188 EVPGMEH 194

RESULT 13  
 G84353  
 hypothetical protein Vng2034h [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: G84353  
 R:McN, W.V.; Kennedy, S.P.; Mahiras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo Jung, K.H.; Alam, M.; Freltas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: G84353  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-274 <STO>  
 A:Cross-references: GB:AE004437; NID:q10581460; PIDN:AAG20195.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG2034H

Query Match  
 Best Local Similarity 79.5%; Score 31; DB 2; Length 274;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 |||||  
 Db 117 EAPGSOH 123

RESULT 14  
 H75446  
 (S)-2-hydroxy-acid oxidase - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: H75446  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Yamathayan, J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: H75446  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-353 <WHI>  
 A:Cross-references: GB:AE001954; GB:AE000513; NID:q6458751; PIDN:AAF10604.1; PID:q645875  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRI031  
 A:Map position: 1  
 C:Superfamily: (s)-2-hydroxy-acid oxidase; (s)-2-hydroxy-acid oxidase homology

F:3-297/Domain: (s)-2-hydroxy-acid oxidase homology <2HY>

Query Match  
 Best Local Similarity 79.5%; Score 31; DB 2; Length 353;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
 |||||  
 Db 169 VPGSEH 194

RESULT 15  
 D81715  
 conserved hypothetical protein TC0328 [imported] - Chlamydia muridarum (strain N1g9)  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Morn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: D81715  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: D81715  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-374 <TEN>  
 A:Cross-references: GB:AE002300; GB:AE002160; NID:q7190365; PIDN:AAF39192.1; PID:q719  
 A:Experimental source: strain N1g9 (Morn)  
 C:Genetics:  
 A:Gene: TC0328

Query Match  
 Best Local Similarity 79.5%; Score 31; DB 2; Length 374;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
 |||||  
 Db 327 LPGSOH 332

Search completed: October 24, 2002, 15:23:16  
 Job time : 8.39344 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:15:42 : Search time 2.7541 Seconds  
(without alignments)  
98.412 Million cell updates/sec

Title: US-09-786-648-2  
Sequence: 1 EYPGSOH 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	124	CHTB_VIBCH	P01556 vibrio chol
2	39	100.0	124	ELBH_ECOLI	P13810 escherichia
3	39	100.0	124	ELBH_ECOLI	P32890 escherichia
4	32	82.1	455	GUDH_BACSU	P42238 bacillus su
5	32	82.1	534	YOGI_CABEL	P34610 caenorhabdi
6	32	82.1	666	PD14_MOUSE	Q92183 mus musculu
7	31	79.5	765	SIM1_MOUSE	Q61045 mus musculu
8	31	79.5	137	YH96_MYCLE	P13733 mycobacteri
9	31	79.5	219	CIDA_HUMAN	O60543 homo sapien
10	31	79.5	240	YB5A_THEMA	P58009 thermotoga
11	31	79.5	273	DLX3_NOTVI	P53770 notophthalp
12	31	79.5	274	DLX3_PLEMA	Q91284 pleurodeles
13	31	79.5	280	DLX3_AMBME	Q90229 ambystoma m
14	31	79.5	414	COT2_BOVIN	O91177 bos tauris
15	31	79.5	414	COT2_HUMAN	P24468 homo sapien
16	31	79.5	414	COT2_MOUSE	P43135 mus musculu
17	31	79.5	414	COT2_RAT	O09010 rattus norv
18	31	79.5	432	AROC_NECR	Q12640 mus musculu
19	31	79.5	459	IL7R_MOUSE	P16872 mus musculu
20	31	79.5	710	IRAI_MOUSE	O62406 mus musculu
21	31	79.5	743	BGAL_THERT	P77989 thermococci
22	30	76.9	182	KPTA_PSEAE	O91778 pseudomonas
23	30	76.9	336	RPOA_THEMA	O94112 thermotoga
24	30	76.9	376	AROC_YEAST	P28777 saccharomyc
25	30	76.9	459	IL7R_HUMAN	P16871 homo sapien
26	30	76.9	500	GABT_HUMAN	P80404 homo sapien
27	30	76.9	573	A37C_DROME	P18487 drosophila
28	30	76.9	504	AMH2_HUMAN	O16671 homo sapien
29	30	76.9	616	RFX5_HUMAN	P48382 homo sapien
30	30	76.9	622	SR68_CABEL	Q20822 caenorhabdi
31	30	76.9	766	SIM1_HUMAN	P81133 homo sapien
32	30	76.9	775	THI1_SCHPO	P36598 schizosacch
33	30	76.9	805	AHR_MOUSE	P30561 mus musculu

34	30	76.9	808	PLD_TOBAC	P93400 nicotiana t
35	30	76.9	853	AHR_RAT	P41738 rattus norv
36	30	76.9	855	ST14_HUMAN	O94576 homo sapien
37	30	76.9	954	DRP2_HUMAN	O13474 homo sapien
38	30	76.9	1286	PATC_DROME	P18502 drosophila
39	30	76.9	1538	LHR_ECOLI	P30015 escherichia
40	30	76.9	1638	BRX_DROME	P25439 drosophila
41	30	76.9	1808	TENA_CHICK	P10039 gallus gall
42	29	74.4	277	2DKG_CORSP	P15339 corynabacte
43	29	74.4	395	INX3_DROME	O9vas7 drosophila
44	29	74.4	404	V4XM_RHIN	P55705 rhizobium s
45	29	74.4	586	RPOO_BMYVF	P09507 beet wester

## ALIGNMENTS

RESULT 1  
CHTB\_VIBCH STANDARD; PRT; 124 AA.  
AC P01556; Q9JQ02;  
DT 21-JUL-1986 (Rel. 01, Created)  
DR 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cholera enterotoxin, beta chain precursor.  
GN CTXB OR TOXB OR VC1456.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lockman H., Kaper J.B.,  
RT "Nucleotide sequence analysis of the A2 and B subunits of Vibrio  
RT cholerae enterotoxin."  
RL J. Biol. Chem. 258:13722-13726(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA STRAIN=EL TOR 2125;  
RC MEDLINE=84061784; PubMed=6315707;  
RA Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F.,  
RT de Wilde M.;  
RT "Cholera toxin genes: nucleotide sequence, deletion analysis and  
RT vaccine development."  
RL Nature 306:351-357(1983).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Dams E., de Wolf M., Dierick W.;  
RC STRAIN=EL TOR 2125;  
RA Submitted (MAY-1991) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA STRAIN=4260B / SEROTYPE O139;  
RC MEDLINE=94237453; PubMed=8181723;  
RA Lebens M., Holmgren J.;  
RT "Structure and arrangement of the cholera toxin genes in Vibrio  
RT cholerae O139."  
RL FEWS Microbiol. Lett. 117:197-202(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA STRAIN=1854 / O139-BENGAL;  
RC Yamanoto K., Do V.G.R.F., Xu M., Iida T., Miwatani T., Albert M.J.,  
RA Honda T.;  
RT Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA STRAIN=EL TOR N16961 / SEROTYPE O1;  
RC MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermlaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 RN [7]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005537; PubMed=903363;  
 RA Kurosky A., Markel D.E., Peterson J.W.;  
 RT "Covalent structure of the beta chain of cholera enterotoxin.";  
 RL J. Biol. Chem. 252:7257-7264(1977).  
 RN [8]  
 RP SEQUENCE OF 22-124.  
 RX MEDLINE=78005536; PubMed=903362;  
 RA Lai C.-Y.;  
 RT "Determination of the primary structure of cholera toxin B subunit.";  
 RL J. Biol. Chem. 252:7249-7256(1977).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=94272319; PubMed=8003954;  
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Hoir C., Martini J.A.,  
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
 RT pentasaccharide.";  
 RL Protein Sci. 3:166-175(1994).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE=95387394; PubMed=7658472;  
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,  
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:  
 RT choleragenoid.";  
 RL J. Mol. Biol. 251:550-562(1995).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RC STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;  
 RX MEDLINE=97376625; PubMed=9232653;  
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,  
 RT "Structural studies of receptor binding by cholera toxin mutants.";  
 RL Protein Sci. 6:1516-1528(1997).  
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
 CC BINDING TO CELL MEMBRANES.  
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN  
 CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
 CC DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
 CC 6 BETA CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, X00171; CAA24996.1; -;  
 DR EMBL, K01170; AAA27573.1; -;  
 DR EMBL, D3A0053; BAA06291.1; -;  
 DR EMBL, X58786; CAA41593.1; -;  
 DR EMBL, X76390; CAA53973.1; -;  
 DR EMBL, X76391; CAA53976.1; -;  
 DR EMBL, AE004224; AAF94613.1; -;  
 DR PIR, A01819; XVCB.  
 DR PIR, A05130; A05130.  
 DR PIR, S14624; S14624.  
 DR PDB, 2CHB; 03-DEC-97.  
 DR PDB, 3CHB; 12-AUG-98.  
 DR PDB, 1CHQ; 08-MAR-96.  
 DR PDB, 1FGB; 23-DEC-96.  
 DR PDB, 1XPB; 01-APR-97.  
 DR PDB, 1XTC; 01-AUG-96.

DR PDB, 1CT1; 15-OCT-97.  
 DR TIGR; VC1456; -;  
 DR InterPro; IPR001835; Enterotoxin\_B.  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 DR Prodom; PD012805; Enterotoxin\_B; 1.  
 KW Membrane; Enterotoxin; Signal; 3D-structure; Complete proteome.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124  
 FT DISULFID 30 107  
 FT CONFLICT 33 33  
 FT CONFLICT 39 39  
 FT CONFLICT 43 43  
 FT CONFLICT 68 68  
 FT CONFLICT 70 70  
 FT CONFLICT 75 75  
 FT CONFLICT 91 91  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT TURN 34 35  
 FT STRAND 36 44  
 FT STRAND 47 51  
 FT TURN 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT TURN 80 99  
 FT HELIX 100 100  
 FT TURN 102 109  
 FT STRAND 115 123  
 SQ SEQUENCE 124 AA; 13957 MW; 9AA3933EAE8E3BDF CRC64;  
 Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best local similarity 100.0%; Pred. No. 0.23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EYVGSQH 7  
 DB 72 EYVGSQH 78  
 ID EYVGSQH 7  
 ID EYVGSQH 78  
 AC P13811;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE heat-labile enterotoxin B chain precursor (LT-B, human) (LTH-B).  
 GN EYVGSQH OR LTPB.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H74-114;  
 RX MEDLINE=85156481; PubMed=3884513;  
 RA Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT infect. Immun. 48:73-77(1985).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE H10407;  
 RX MEDLINE=83114628; PubMed=6759877;  
 RA Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
 RT "Overlapping genes in the heat-labile enterotoxin operon originating  
 RT from Escherichia coli human strain.";  
 RL Mol. Gen. Genet. 188:356-359(1982).  
 RN [3]  
 RN SEQUENCE FROM N.A.

```

RC STRAIN-ISOLATE H10407;
RX MEDLINE=93252225; PubMed=8486242;
RA Inoue T., Tsuji T., Koto M., Imanura S., Miyama A.;
RT "Amino acid sequence of heat-labile enterotoxin from chicken
RT enterotoxigenic Escherichia coli is identical to that of human strain
RT H 10407."
RL FEWS Microbiol. Lett. 108:157-161(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ETEC LT 87;
RA Germani Y., Desperrier J.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=95349400; PubMed=7623669;
RA Domenighini M., Piza M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae."
RL Mol. Microbiol. 15:1165-1167(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).
RX MEDLINE=99185101; PubMed=10085117;
RA Matkovic-Calogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,
RA Tossi A., Palu G., Zanotti G.;
RT "Crystal structure of the B subunit of Escherichia coli heat-labile
RT enterotoxin carrying peptides with anti-herpes simplex virus type 1
RT activity."
RL J. Biol. Chem. 274:8764-8769(1999).
CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17874; AAA98064.1; -
DR EMBL: J01646; AAC02982.1; -
DR EMBL: S60731; AAC60441.1; -
DR EMBL: X83966; CAA58800.1; -
DR PDB: 1LTR; 2J-MAR-99.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B.1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR PRODOM: PD012805; Enterotoxin_B.1.
KW Enterotoxin; Signal; 3p-structure.
FT SIGNAL 1 21
FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
FT DISULFID 30 107
FT VARIANT 6 6 F -> C (IN ISOLATE H10407).
FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).
FT VARIANT 34 34 H -> R (IN ISOLATE H10407).
SQ SEQUENCE 124 AA; 14027 MW; E3FF7C7B9D3BC47 CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVPSQSH 7
Db 72 EVPSQSH 78

```

```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DR Heat-labile enterotoxin B chain precursor (LT-B, porcine) (LTB-B).
GN LTB or LTPB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE P307;
RX MEDLINE=81074965; PubMed=7003397;
RA Dallas W.S., Falkow S.;
RT "Amino acid sequence homology between cholera toxin and Escherichia
RT coli heat-labile toxin."
RL Nature 288:499-501(1980).
RN [2]
RP REVISIONS TO 28 AND 64.
RC STRAIN-ISOLATE P307;
RX MEDLINE=85156481; PubMed=3884513;
RA Leong J., Vinal A.C., Dallas W.S.;
RT "Nucleotide sequence comparison between heat-labile toxin B-subunit
RT cistrons from Escherichia coli of human and porcine origin."
RL Infect. Immun. 48:73-77(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE PC86;
RX MEDLINE=87137303; PubMed=3546273;
RA Yamamoto T., Gojohori T., Yokota T.;
RT "Evolutionary origin of pathogenic determinants in enterotoxigenic
RT Escherichia coli and Vibrio cholerae O1."
RL J. Bacteriol. 169:1352-1357(1987).
RN [4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=87280041; PubMed=3301830;
RA Ibrahim I., Gentz R.;
RT "A functional interaction between the signal peptide and the
RT translation apparatus is detected by the use of a single point
RT mutation which blocks translocation across mammalian endoplasmic
RT reticulum."
RL J. Biol. Chem. 262:10189-10194(1987).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX MEDLINE=93240541; PubMed=8478941;
RA Sliwa T.K., van Zanten B.A.M., Dauter Z., Hol W.G.J.;
RT "Refined structure of Escherichia coli heat-labile enterotoxin, a
RT close relative of cholera toxin."
RL J. Mol. Biol. 230:890-918(1993).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91238966; PubMed=2034287;
RA Sliwa T.K., Pronk S.E., Kalk K.H., Martina E.S., van Zanten B.A.M.,
RA Wilcholt B., Hol W.G.J.;
RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin
RT from E. coli."
RL Nature 351:371-377(1991).
RN [7]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=95349400; PubMed=7623669;
RA Domenighini M., Piza M., Jobling M.G., Holmes R.K., Rappuoli R.;
RT "Identification of errors among database sequence entries and
RT comparison of correct amino acid sequences for the heat-labile
RT enterotoxins of Escherichia coli and Vibrio cholerae."
RL Mol. Microbiol. 15:1165-1167(1995).
CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC or send an email to license@lsb-sib.ch).
-----
CC EMBL; M17873; AAA98065.1; -
CC DR EMBL; M15363; AAA24792.1; -
CC DR EMBL; M17101; AAA23973.1; -
CC PIR; A01820; OLECB.
CC DR PIR; B26946; OLECEB.
CC DR PDB; 1LTA; 31-JAN-94.
CC DR PDB; 1LTB; 31-JAN-94.
CC DR PDB; 1LTG; 15-SEP-95.
CC DR PDB; 1LTI; 17-AUG-96.
CC DR PDB; 1LTS; 31-JAN-94.
CC DR PDB; 1LTT; 31-JAN-94.
CC DR PDB; 1LTS; 07-JUL-97.
CC DR PDB; 1LTA; 16-JUN-97.
CC DR PDB; 1LTS; 03-DEC-97.
CC DR PDB; 1LTV; 03-DEC-97.
CC DR PDB; 1HTU; 20-APR-95.
CC DR InterPro; IPR001835; Enterotoxin_B.
CC DR Pfam; PF01376; Enterotoxin_B; 1.
CC DR PRINTS; PR00772; ENTEROTOXINB.
CC DR ProDom; PD012805; Enterotoxin_B; 1.
CC DR Enterotoxin; Signal; 3D-structure.
CC FT SIGNAL 1 21
CC FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.
CC FT DISULFD 30 107
CC FT HELIX 26 30
CC FT TURN 31 32
CC FT STRAND 36 43
CC FT STRAND 47 51
CC FT TURN 54 55
CC FT STRAND 58 62
CC FT TURN 64 65
CC FT STRAND 68 71
CC FT TURN 76 77
CC FT HELIX 80 98
CC FT TURN 99 100
CC FT STRAND 103 109
CC FT STRAND 115 123
CC SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE8395EA70D CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. NO. 0.23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
Db 72 EVPGSQH 78

RESULT 4
GUDH_BACSU
ID GUDH_BACSU STANDARD; PRT: 455 AA.
AC P42238;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable gluconate dehydratase (EC 4.2.1.40) (GDH) (GLUCD).
GN GUDH
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RX MEDLINE=95219079; PubMed=7704254;
RA "Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24
RL Microbiology 141:269-275(1995).
-1- FUNCTION:CATALYZES THE DEHYDRATION OF GLUCONATE TO 5-KETO-4-

```

```

CC CC XEROXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY) :
CC CC CATABOLYTIC ACTIVITY: D-glucarate - 5-dehydro-4-deoxy-D-glucarate +
CC CC H(2)O.
CC CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.
CC CC -1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE
CC CC LACTONIZING ENZYME FAMILY. GUCC SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D30808; BAA06470.1; -.
CC CC EMBL; D99105; CAB12043.1; -.
CC CC HSSP; P42206; IBOG.
CC CC Subtilist; BG11161; gudd.
CC CC InterPro; IPR001354; MR_MLE.
CC CC Pfam; PF01188; MR_MLE.1.
CC CC Pfam; PF02746; MR_MLE_N.1.
CC CC Lyase; Complete proteome.
CC CC SEQUENCE 455 AA; 50782 MW; 32384860076982A CRC64;
CC CC -----
SQ
Query Match 82.1%; Score 32; DB 1; Length 455;
Best Local Similarity 71.4%; Pred. NO. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 EVPGSGH 7
Db 64 EVPGGEH 70
-----
RESULT 5
YOG1_CAEEL STANDARD; PRT; 534 AA.
ID YOG1_CAEEL
AC P34610;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative serine protease ZK112.1 precursor (EC 3.4.-.-).
EN ZK112.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloiderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN R1
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Harkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lathelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Stulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:37-38(1994).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S28 (SERINE PROTEASE).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----

```



```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: L14324: AAA28181.1; -.
DR PIR: S44886; S44886.
DR WormPep: ZK112.1; CE00372.
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000379; Est_lip_thioest_actstle.
DR Pfam: PF00561; abhydrolase: 1.
KW Hypothetical protein: Hydrolase; Serine protease; Signal.
FT SIGNAL 1 17
FT CHAIN 1 17
FT ACT_SITE 18 534
FT ACT_SITE 177 177
FT ACT_SITE 420 420
FT ACT_SITE 448 448
FT CARBOHYD 69 69
FT CARBOHYD 107 107
FT CARBOHYD 126 126
FT CARBOHYD 240 240
FT CARBOHYD 244 244
FT CARBOHYD 257 257
FT CARBOHYD 271 271
FT CARBOHYD 288 288
FT CARBOHYD 316 316
SQ SEQUENCE 534 AA; 59221 MW; C43F0104B42E4630 CRC64;

Query Match
Best Local Similarity 82.1%; Score 32; DB 1; Length 534;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
Db 442 EIPGSAH 448

RESULT 6
SIML_MOUSE
ID PD14_MOUSE STANDARD; PRT; 666 AA.
AC 092183;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein-arginine deiminase type IV (EC 3.5.3.15) (Peptidylarginine
DE deiminase IV).
GN PD14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epidermis;
RX MEDLINE=9192810; PubMed=10092850;
RA Rusc A.A., Ikejiri Y., Ono H., Yonekawa T., Shiraiwa M., Kawada A.,
RA Takahara H.;
RT "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type
RT I, type III and type IV, and the expression pattern of type I in
RT mouse";
RL Eur. J. Biochem. 259:660-669(1999).
CC -1- FUNCTION: CATALYZES THE DEIMINATION OF ARGININE RESIDUES OF
CC PROTEINS.
CC -1- CATALYTIC ACTIVITY: Protein L-arginine + H(2)O = protein L-
CC citrulline + NH(3)
CC -1- COFACTOR: REQUIRES CALCIUM IONS.
CC -1- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND TO ARGININE
CC DEIMINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
```

```

CC -----
DR EMBL: AB013850; BAA34246.1; -.
DR MGD: MGI:1338698; P814.
DR InterPro: IPR004303; PAD.
DR Pfam: PF03068; PAD; 1.
KW Hydrolase; Calcium-binding; Multigene family.
FT CA_BIND 505 516
FT SEQUENCE 666 AA; 74476 MW; 70FAFE7E232D34A CRC64;

Query Match
Best Local Similarity 82.1%; Score 32; DB 1; Length 666;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSQH 7
Db 239 EIPGSAH 245

RESULT 7
SIML_MOUSE
ID SIML_MOUSE STANDARD; PRT; 765 AA.
AC 061045; P70183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 1 (mSIM1).
GN SIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER;
RX MEDLINE=97020303; PubMed=8812055;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RT "Expression patterns of two murine homologs of Drosophila
RT single-minded suggest possible roles in embryonic patterning and in
RT the pathogenesis of Down syndrome.";
RL Mol. Cell. Neurosci. 7:1-16(1996).
RN [2]
RP ERRATUM.
RX MEDLINE=97029422; PubMed=8875433;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RL Mol. Cell. Neurosci. 7:519-519(1996).
RN [3]
RP REVISIONS TO C-TERMINUS.
RX MEDLINE=97343329; PubMed=919934;
RA Fan C.-M.;
RT Unpublished results, cited by:
RL Chrest R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,
RL Wang Y., Shimizu N., Antonarakis S.E.;
RL Genome Res. 7:615-624(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=96413339; PubMed=8927054;
RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,
RA Saijoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;
RT "Two new members of the murine Sim gene family are transcriptional
RT repressors and show different expression patterns during mouse
RT embryogenesis.";
RL Mol. Cell. Biol. 16:5865-5875(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Hosoya T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
```

RP SUBUNIT.  
 RA MEDLINE-97172525; PubMed-9020169;  
 RX Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;  
 RT "Two murine homologs of the Drosophila single-minded protein that  
 RT interact with the mouse aryl hydrocarbon receptor nuclear  
 RT translocator protein.";  
 RL J. Biol. Chem. 272:4451-4457(1997).  
 CC -1- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC -1- TISSUE SPECIFICITY: DETECTED IN LUNG, SKELETAL MUSCLE AND KIDNEY.  
 CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING  
 CC SOMITES, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING  
 CC KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; U40575; AAA91201.1; ALT\_SEQ.  
 CC EMBL; D79209; BAA11467.1; -.  
 CC EMBL; AB013491; BAA28270.1; -.  
 CC EMBL; AB013484; BAA28270.1; JOINED.  
 CC EMBL; AB013485; BAA28270.1; JOINED.  
 CC EMBL; AB013486; BAA28270.1; JOINED.  
 CC EMBL; AB013487; BAA28270.1; JOINED.  
 CC EMBL; AB013488; BAA28270.1; JOINED.  
 CC EMBL; AB013489; BAA28270.1; JOINED.  
 CC EMBL; AB013490; BAA28270.1; JOINED.  
 CC MGD; MGI:96306; Sim1.  
 CC InterPro: IPR003015; HLH\_Myc.  
 CC InterPro: IPR001092; HLH\_dlm.  
 CC InterPro: IPR001067; NuclearLocator.  
 CC InterPro: IPR001610; PAC.  
 CC InterPro: IPR000014; PAS.  
 CC Pfam; PF00785; PAC; 1.  
 CC Pfam; PF00989; PAS; 2.  
 CC PRINTS; PR00785; NCTRNLOCATR.  
 CC SMART; SM00353; HLH; 1.  
 CC SMART; SM00086; PAC; 1.  
 CC SMART; SM00091; PAS; 2.  
 CC PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 CC PROSITE; PS50112; PAS; 2.  
 CC Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 CC Transcription regulation; DNA-binding.  
 CC KW DNA BIND  
 CC FT DOMAIN 1 13 BASIC DOMAIN.  
 CC FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 CC FT DOMAIN 77 147 PAS 1.  
 CC FT DOMAIN 218 288 PAS 2.  
 CC FT CONFLICT 133 133 H -> L (IN REF. 1).  
 CC FT CONFLICT 176 176 MISSING (IN REF. 1).  
 CC FT CONFLICT 322 322 P -> R (IN REF. 1).  
 CC FT CONFLICT 480 480 A -> P (IN REF. 1).  
 CC FT CONFLICT 537 537 D -> S (IN REF. 1).  
 CC SEQUENCE 765 AA; 85540 MW; B1A7FDAB8578CD17 CRC64;

Query Match 82.1%; Score 32; DB 1; Length 765;  
 Best Local Similarity 85.7%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EVPGSOH 7  
 Db 423 ERPGSOH 429

RESULT 8  
 YH96\_MYCLE  
 ID YH96\_MYCLE STANDARD; PRT; 137 AA.  
 AC P13733;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 15.2 kDa protein M11796.  
 GN M11796.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxId=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-88088878; PubMed-2447183;  
 RA Booth R.J., Harris D.P., Love J.M., Watson J.D.;  
 RT "Antigenic proteins of Mycobacterium leprae. Complete sequence of the  
 RT gene for the 18-kDa protein.";  
 RL J. Immunol. 140:597-601(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TN:  
 RX MEDLINE-21128732; PubMed-11234002;  
 RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Dunning P.R., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares K.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; M19058; AAA88230.1; -.  
 CC EMBL; AL583923; CAC30749.1; -.  
 CC PIR; B27586; B27586.  
 CC Leproma; M11796; -.  
 CC KW Hypothetical protein; Complete proteome.  
 CC SEQUENCE 137 AA; 15195 MW; 633B6B8F3FDAD0B CRC64;

Query Match 79.5%; Score 31; DB 1; Length 137;  
 Best Local Similarity 83.3%; Pred. No. 12;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 VPGSOH 7  
 Db 127 VPGNOH 132

RESULT 9  
 CIDA\_HUMAN  
 ID CIDA\_HUMAN STANDARD; PRT; 219 AA.  
 AC O60543;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cell death activator CIDE-A (Cell death-inducing DFFA-like effector  
 DE A).  
 GN CIDEA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98232498; PubMed=9564035;
RA Inohara N., Koseki T., Chen S., Wu X., Nunez G.;
RT "CIDE", a novel family of cell death activators with homology to the 45
RT kDa subunit of the DNA fragmentation factor."
RL EMO J. 17:2526-2533(1998).
CC -1- FUNCTION: ACTIVATES APOPTOSIS.
CC -1- SUBUNIT: INHIBITED BY DFFB.
CC -1- SIMILARITY: CONTAINS 1 CIDE-N DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AF041378; AAC34987.1; -.
DR HSSP: Q90HD4; ID4B.
DR MIM: 604440; -.
DR InterPro: IPR003508; CAD.
DR Pfam: PF02017; CIDE-N; 1.
DR SMART: SM00266; CAD; 1.
DR Apoptosis.
KW DOMAIN.
FT 33 110 CIDE-N.
SQ SEQUENCE 219 AA; 24686 MW; 05F704823CE71C0E CRC64;

Query Match 79.5%; Score 31; DB 1; Length 219;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSOH 7
Db 109 MPGSOH 114

RESULT 10
YB5A_THEMEA STANDARD; PRT; 240 AA.
ID YB5A_THEMEA
AC P58009;
DE 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TM1158.1.
GN TM1158.1.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterlinden T.R., Malek J.A., Ianher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
RN [2]
RP IDENTIFICATION.
RA Medigue C., Bocs S.;
RL Unpublished observations (Apr-2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AE001773; -; NOT_ANNOTATED_CDS.
DR TIGR: TM1158.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 240 AA; 27773 MW; 1BDF66C1C8BD2700 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 240;
Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSEH 7
Db 226 VPGSEH 231

RESULT 11
DLX3_NOTVT STANDARD; PRT; 273 AA.
ID DLX3_NOTVT
AC P53770;
DE 01-OCT-1996 (Rel. 34, Created)
DE 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Homeobox protein DLX-3 (Box-4) (NKHBOX-4).
GN DLX3 OR BOX4.
OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;
OC Notophthalmus.
OX NCBI_TaxID=8316;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RX MEDLINE=93050784; PubMed=1358728;
RA Beauchemin M., Savard P.;
RT "Two distal-less related homeobox-containing genes expressed in
RT regeneration blastemas of the newt."
RL Dev. Biol. 154:55-65(1992).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN REGENERATION BLASTEMAS, DETECTED
CC IN FORELIMBS, HINDLIMBS, THE TAIL, FLANK, AND BRAIN AS WELL AS IN
CC LIMB AND TAIL BLASTEMAS.
CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: X63531; CA945094.1; -.
DR HSSP: P22808; INK3.
DR InterPro: IPR000047; HTH_repressr.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX-1; 1.
DR PROSITE: PS00071; HOMEBOX-2; 1.
FW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 125 184 HOMEBOX
SQ SEQUENCE 273 AA; 30654 MW; B356D01233061F2F CRC64;

Query Match 79.5%; Score 31; DB 1; Length 273;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 EYVGSOH 7
      |||| :|
Db      188 EYVGMH 194

RESULT 12
DLX3_PLEWA
ID      DLX3_PLEWA      STANDARD:      PRT:      274 AA.
AC      Q91284;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      Distal-LESS like protein DLX-3.
GN      DLX3.
OS      Pleurodeles waltl11 (Iberian ribbed newt).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;
OC      Pleurodeles.
OX      NCBI_TaxID=8319;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96391186; PubMed=8798159;
RA      Nicolas S., Massacrier A., Caubit X., Cau P., le Parco Y.;
RT      "A Distal-less like gene is induced in the regenerating central
RT      nervous system of the urodele Pleurodeles waltl.";
RL      Mech. Dev. 56:209-220(1996).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC      -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE REGENERATING SPINAL CORD BUT
CC      NOT IN THE ADULT ONE.
CC      -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEOBOX
CC      PROTEINS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; U49645; AAA98645.1; -.
DR      HSSP; P22808; INK3.
DR      InterPro; IPR000047; HTH_repressor.
DR      InterPro; IPR001356; Homeobox.
DR      Pfam; PF00046; homeobox; 1.
DR      PRINTS; PR00024; HOMEOBOX.
DR      PRINTS; PR00031; HTHREPRESSR.
DR      SMART; SM00389; HOX; 1.
DR      PROSITE; PS00027; HOMEOBOX_1; 1.
DR      PROSITE; PS50071; HOMEOBOX_2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT      DNA_BIND 126 185 HOMEOBOX.
SQ      SEQUENCE 274 AA; 30607 MW; 94148553BA808C8 CRC64;

Query Match      79.5%; Score 31; DB 1; Length 274;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      1 EYVGSOH 7
      |||| :|
Db      188 EYVGMH 195

RESULT 13
DLX3_AMBME
ID      DLX3_AMBME      STANDARD:      PRT:      280 AA.
AC      Q90229;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Homeobox protein DLX-3.
GN      DLX-3.

```

```

OS      Ambystoma mexicanum (Axolotl).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;
OC      Ambystoma.
OX      NCBI_TaxID=8296;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97108743; PubMed=8951064;
RA      Mullen L.M., Bryant S.V., Torok M.A., Blumberg B., Gardiner D.M.;
RT      "Nerve dependency of regeneration: the role of Distal-less and FGF
RT      signaling in amphibian limb regeneration.";
RL      Development 122:3487-3497(1996).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC      -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEOBOX
CC      PROTEINS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; U59480; ABA9668.1; -.
DR      HSSP; P22808; INK3.
DR      InterPro; IPR000047; HTH_repressor.
DR      InterPro; IPR001356; Homeobox.
DR      Pfam; PF00046; homeobox; 1.
DR      PRINTS; PR00024; HOMEOBOX.
DR      PRINTS; PR00031; HTHREPRESSR.
DR      SMART; SM00389; HOX; 1.
DR      PROSITE; PS00027; HOMEOBOX_1; 1.
DR      PROSITE; PS50071; HOMEOBOX_2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT      DNA_BIND 126 185 HOMEOBOX.
SQ      SEQUENCE 280 AA; 31200 MW; 0F8183097ABAC791 CRC64;

Query Match      79.5%; Score 31; DB 1; Length 280;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      1 EYVGSOH 7
      |||| :|
Db      189 EYVGMH 195

RESULT 14
COT2_BOVIN
ID      COT2_BOVIN      STANDARD:      PRT:      414 AA.
AC      Q9TTR7;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Coup transcription factor 2 (COT2-TF2) (COT2-TF II).
GN      NR2F2 OR TFCDP2.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Corpus luteum;
RA      Walther N.;
RL      Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
CC      -1- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
CC      -1- BINDS TO DNA SITE A (BY SIMILARITY).
CC      -1- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC      NR2 SUBFAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ249441; CAB55624.1; -
DR HSSP; P19793; 2NLL.
DR InterPro; IPR000536; Hormone_rec_1lg.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec. 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR Transcription regulation; DNA-binding; Nuclear protein;
DR Zinc-finger.
DR DOMAIN; 71 75 POLY-GLN.
DR DNA_BIND; 79 144 NUCLEAR_RECEPTOR-TYPE.
DR ZN_FING; 79 99 C4-TYPE.
DR ZN_FING; 115 139 C4-TYPE.
DR DOMAIN; 337 414 IMPORTANT FOR DIMERIZATION (BY
DR FT SIMILARITY).
DR SQ SEQUENCE 414 AA; 45557 MW; C24CB2023C8A27F57 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSQ 6
    |||||
Db 14 EYPSQ 19

RESULT 15
COT2_HUMAN STANDARD; PRT; 414 AA.
ID COT2_HUMAN
AC P24468; 003754;
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE COUP transcription factor 2 (COUP-TF2) (Apolipoprotein AI
DE regulatory protein-1) (ARP-1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91118002; PubMed=1899293;
RA Ladias J.A.A., Karathanasis S.K.;
RT "Regulation of the apolipoprotein AI gene by ARP-1, a novel member of
RT the steroid receptor superfamily.";
RL Science 251:561-565(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Speckmayer R.W.M., Paulweber B., Sandhofer F.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-351 FROM N.A.
RX MEDLINE=92314709; PubMed=1820218;
RA Wang L.H., Ing N.H., Tsai S.Y., O'Malley B.W., Tsai M.J.;
RT "The COUP-TFs compose a family of functionally related transcription
RT factors.";
RL Gene Expr. 1:207-216(1991).
CC -!- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
CC -!- BINDS TO DNA SITE A.

```

```

CC -!- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M64497; AA86429.1; -
DR EMBL; U60477; AA809475.1; -
DR EMBL; M62760; AAA21479.1; -
DR PIR; A37133; A37133.
DR HSSP; P19793; 2NLL.
DR TRANSFAC; T00045; -
DR MIM; 107773; -
DR InterPro; IPR000536; Hormone_rec_1lg.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec. 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR Transcription regulation; DNA-binding; Nuclear protein;
DR Zinc-finger.
DR DOMAIN; 71 75 POLY-GLN.
DR DNA_BIND; 79 144 NUCLEAR_RECEPTOR-TYPE.
DR ZN_FING; 79 99 C4-TYPE.
DR ZN_FING; 115 139 C4-TYPE.
DR DOMAIN; 337 414 IMPORTANT FOR DIMERIZATION.
DR SQ SEQUENCE 414 AA; 45571 MW; C24CB2E8C8A27E8C CRC64;

Query Match 79.5%; Score 31; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYPSQ 6
    |||||
Db 14 EYPSQ 19

```

Search completed: October 24, 2002, 15:20:59  
 Job time : 5.7541 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen, Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 15:16:03 : Search time 8.72131 Seconds  
(without alignments)  
138.851 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EWPGSQH 7

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.podent:\*  
13: sp.virus:\*  
14: sp.vertibrate:\*  
15: sp.unclassified:\*  
16: sp.virus:\*  
17: sp.bacteriap:\*  
17: sp.archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	100.0	103	2	Q9R646
2	39	100.0	124	2	Q57193
3	39	100.0	124	2	Q9R615
4	39	100.0	124	2	Q56635
5	39	100.0	124	2	Q93V32
6	39	100.0	124	9	Q94M01
7	35	89.7	565	10	Q22511
8	34	87.2	195	13	Q9W7D3
9	34	87.2	199	13	Q9W7D4
10	34	87.2	201	2	Q9R618
11	33	84.6	192	4	Q9NPO6
12	33	84.6	272	4	Q9UG04
13	33	84.6	367	10	Q80418
14	33	84.6	412	4	Q9Y2B3
15	33	84.6	868	5	Q9VFE2
16	32	82.1	95	10	Q9F018

17	32	82.1	137	16	Q92T78	Q92T78 rhizobium m
18	32	82.1	765	11	Q70284	Q70284 mus musculus
19	31	79.5	95	10	Q9F0J8	Q9F0J8 amarantus
20	31	79.5	122	5	Q9W2V0	Q9W2V0 drosophila
21	31	79.5	134	16	Q92N40	Q92N40 rhizobium m
22	31	79.5	222	11	Q9JL95	Q9JL95 mus musculus
23	31	79.5	233	10	Q94KK7	Q94KK7 arabisidopsis
24	31	79.5	242	2	Q86582	Q86582 streptomyces
25	31	79.5	243	11	Q90Y63	Q90Y63 mus musculus
26	31	79.5	260	10	Q9MA16	Q9MA16 arabidopsis
27	31	79.5	274	17	Q9HNM5	Q9HNM5 halobacteri
28	31	79.5	331	5	Q18391	Q18391 delnoccocus
29	31	79.5	353	16	Q9RVJ7	Q9RVJ7 zymomonas m
30	31	79.5	360	11	Q9PD68	Q9PD68 chlamydia m
31	31	79.5	371	2	Q9RH15	Q9RH15 drosophila
32	31	79.5	374	16	Q9PKY2	Q9PKY2 cryptonec
33	31	79.5	382	5	Q9V7M7	Q9V7M7 cryptonec
34	31	79.5	406	3	Q9XUM8	Q9XUM8 caenorhabdi
35	31	79.5	432	3	Q9P3J3	Q9P3J3 mus musculus
36	31	79.5	459	11	Q9ROC1	Q9ROC1 mus musculus
37	31	79.5	460	10	Q9LIF8	Q9LIF8 arabidopsis
38	31	79.5	472	2	Q93913	Q93913 klebsiella
39	31	79.5	489	5	Q9XUM8	Q9XUM8 caenorhabdi
40	31	79.5	496	11	Q9D0H2	Q9D0H2 mus musculus
41	31	79.5	497	4	Q9BR66	Q9BR66 homo sapien
42	31	79.5	510	4	Q9H846	Q9H846 homo sapien
43	31	79.5	510	4	Q9G6M8	Q9G6M8 homo sapien
44	31	79.5	569	6	Q95UP7	Q95UP7 macaca fasc
45	31	79.5	660	10	Q9AY34	Q9AY34 oryza sativ

## ALIGNMENTS

RESULT 1  
ID Q9R646 PRELIMINARY: PRT: 103 AA.  
AC Q9R646:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CHOLERA-LIKE ENTEROTOXIN B SUBUNIT.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=95303036; PubMed=7783690;  
RA Nakashima K., Eguchi Y., Nakasone N.;  
RT "Characterization of an enterotoxin produced by Vibrio cholerae  
RT 0139."  
RL Microbiol. Immunol. 39:87-94(1995).  
DR HSSP: P01556; IYTC.  
DR InterPro: IPR001835; Enterotoxin\_B.  
DR Pfam: PF01376; Enterotoxin\_B: 1.  
DR PRINTS: PR00772; ENTEROTOXINB.  
DR PRODOM: PD012805; Enterotoxin\_B: 1.  
SQ SEQUENCE 103 AA; 11645 MW; 992A05C077F2F70E CRC64;  
Query Match 100.0%; Score 39; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EWPGSQH 7  
DB 51 EWPGSQH 57  
RESULT 2  
ID Q57193 PRELIMINARY: PRT: 124 AA.  
AC Q57193:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)

```

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA TOXIN B PROTEIN (CTB) PRECURSOR (CTB).
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL STRAIN 569B;
RX MEDLINE=91355224; PubMed=1883840;
RA Dame E., De Wolf M., Dierick W.;
RT "Nucleotide sequence analysis of the CT operon of the Vibrio cholerae
RL classical strain 569B."
RN [2]
RP Biochim. Biophys. Acta 1090:139-141 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Shi C., Cao C., Zhang J., Ma Q.;
RL Chin. Biochem. J. 9:395-399 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL BIOTYPE 569B;
RA Xu L.;
RN [3]
RP Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
RA EMBL: X58785; CAA41591.1; -.
DR EMBL: 025679; AAC34728.1; -.
DR EMBL: A00931; CAA00098.1; -.
DR HSSP: A01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B. 1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR Prodom: PD012805; Enterotoxin_B. 1.
DR Signal.
FT SIGNAL.
KW CHOLERA TOXIN B PROTEIN (CTB).
SQ SEQUENCE 124 AA; 13919 MW; D6BF83FF7924EA3 CRC64;

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
DB 72 EVPGSOH 78

RESULT 3
O9RP15 PRELIMINARY; PRT; 124 AA.
AC O9RP15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA ENTEROTOXIN B SUBUNIT.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KNIH002;
RA Shin H.-J., Park Y.C., Kim Y.C.;
RT "Cloning and nucleotide sequence analysis of the virulence gene
RL cassette from Vibrio cholerae KNIH002 isolated in Korea."
RN [2]
RP Misilmungh Hoil 35:205-210 (1999).
DR EMBL: AF175708; AAD51360.1; -.
DR HSSP: P01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B. 1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR Prodom: PD012805; Enterotoxin_B. 1.
SQ SEQUENCE 124 AA; 13905 MW; 23BF83FF7935B9 CRC64;

```

```

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
DB 72 EVPGSOH 78

RESULT 4
O56635 PRELIMINARY; PRT; 124 AA.
ID O56635;
AC O56635;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CHOLERA TOXIN PRECURSOR.
GN CTXB.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S7;
RA Yamamoto K., Do V.G., Xu M., Iida T., Mawatani T., Albert M.J.,
RA Honda T.;
RT "Comparison of cholera toxin genes (ctxAB) of non-O1 vibrio cholerae
RL strains 854 (O139-bengal) and S7 (O37) from two outbreaks."
DR EMBL: D30052; BAA06289.1; -.
DR HSSP: P01556; 2CHB.
DR InterPro: IPR001835; Enterotoxin_B.
DR Pfam: PF01376; Enterotoxin_B. 1.
DR PRINTS: PR00772; ENTEROTOXINB.
DR Prodom: PD012805; Enterotoxin_B. 1.
DR Signal.
FT SIGNAL.
KW CHOLERA TOXIN B PROTEIN (CTB).
SQ SEQUENCE 124 AA; 13871 MW; 3F87B2F297953179 CRC64;

Query Match 100.0%; Score 39; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
DB 72 EVPGSOH 78

RESULT 5
O93V32 PRELIMINARY; PRT; 124 AA.
ID O93V32;
AC O93V32;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HEAT-LABILE ENTEROTOXIN B SUBUNIT.
GN LTH B SUBUNIT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RA Komase K.;
RT Submitted (Mar-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1032 (ENTEROTOXIGENIC);
RX MEDLINE=95091056; PubMed=7998417;
RA Tamura S., Asanuma H., Tomita T., Komase K., Kawahara K., Danbara H.,
RA Hattori N., Watanabe K., Suzuki Y., Nagamine T., Aizawa C., Oya A.,

```



RA Kurata T.;  
 RT "Escherichia coli heat-labile enterotoxin B subunits supplemented with  
 RT a trace amount of the holotoxin as an adjuvant for nasal influenza  
 RT vaccine.";  
 RL Vaccine 12:1083-1089(1994).  
 DR EMBL: AB011677; BAA25726.1; -  
 SQ SEQUENCE 124 AA; 14028 MW; 5346BD38B32354C2 CRC64;

Query Match 100.0%; Score 39; DB 2; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 |||||  
 DB 72 EVPGSOH 78

## RESULT 6

ID 094M01 PRELIMINARY; PRT; 124 AA.

AC 094M01;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE CTXB.  
 GN CTXB.  
 OS Vibrio phage CTX.  
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxID=141904;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Battacharyya T., Nandy R.K., Nair G.B.;  
 RT "The entire core region of the ctx-phi (ctx-phage) in VCE 232, an  
 RT environmental strain of V. cholerae.";  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF414369; AAL09682.1; -  
 SQ SEQUENCE 124 AA; 13884 MW; 8D4C251591B16891 CRC64;

Query Match 100.0%; Score 39; DB 9; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 |||||  
 DB 72 EVPGSOH 78

## RESULT 7

ID 022511 PRELIMINARY; PRT; 565 AA.

AC 022511;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE GLUTATHIONE REDUCTASE (NADPH) (EC 1.6.4.2) (FRAGMENT).  
 GN GOR.

OS Vitis vinifera (Grape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;  
 OC Vitis.  
 OX NCBI\_TaxID=29760;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV. THOMPSON SEEDLESS (CLONE 2A); TISSUE=FRUIT;  
 RA Cassol T., Adams D.O.;  
 RT "Cloning of a grape glutathione reductase cDNA and analysis of its  
 RT expression during berry development.";  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1-COFACITOR: FAD (BY SIMILARITY).  
 DR EMBL: AF019907; AAB70837.1; -  
 DR HSSP; P06715; IGER.  
 DR InterPro: IPR001327; FAD\_Pyr\_redox.  
 DR InterPro: IPR000103; Pyridine\_redox\_2.

DR InterPro: IPR001100; Pyr\_redox.  
 DR InterPro: IPR004099; Pyr\_redox\_dlm.

DR Pfam; PF000070; Pyr\_redox; 1.  
 DR Pfam; PF02852; Pyr\_redox\_dlm; 1.  
 DR PRINTS; PR00368; FADPNR.  
 DR PRINTS; PR00411; PNDROTAESI.  
 DR PRINTS; PR00469; PNDROTAESI.  
 DR PROSITE; PS00076; PYRIDINE\_REDOX\_1; 1.  
 KW FAD; Flavoprotein; Oxidoreductase.  
 FT NON\_TER 1  
 SQ SEQUENCE 565 AA; 60695 MW; B26113AE09A121DE CRC64;

Query Match 89.7%; Score 35; DB 10; Length 565;  
 Best Local Similarity 71.4%; Pred. No. 25;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 |||||  
 DB 239 EVPGSOH 245

## RESULT 8

ID 09W7D3 PRELIMINARY; PRT; 195 AA.

AC 09W7D3;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOTHEICAL 21.4 KDA PROTEIN.  
 OS Oryzias latipes (Medaka fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.  
 OX NCBI\_TaxID=8090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-ORANGE-RED;  
 RX MEDLINE=20070874; Pubmed=10602271;  
 RA Kanamori A.;  
 RT "Systematic identification of genes expressed during early oogenesis  
 RT in medaka.";  
 RL Mol. Reprod. Dev. 55:31-36(2000).  
 DR EMBL: AF128618; AAD38915.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 195 AA; 21446 MW; B90682D695729E88 CRC64;

Query Match 87.2%; Score 34; DB 13; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPGSOH 7  
 |||||  
 DB 111 VPGSOH 116

## RESULT 9

ID 09W7D4 PRELIMINARY; PRT; 199 AA.

AC 09W7D4;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOTHEICAL 21.7 KDA PROTEIN.  
 OS Oryzias latipes (Medaka fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.  
 OX NCBI\_TaxID=8090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-ORANGE-RED;

RX MEDLINE=20070874; PubMed=10602271;  
 RA Kanamori A.;  
 RT "Systematic identification of genes expressed during early oogenesis  
 in medaka.";  
 RL Mol. Reprod. Dev. 55:31-36(2000).  
 DR EMBL: AF128817; AAD38914.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 199 AA; 21726 MW; E9827C05451B15CD CRC64;

Query Match 87.2%; Score 34; DB 13; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 EYVGSQH 7  
 DB 115 EYVGSQH 120

## RESULT 10

O9RJZ8 PRELIMINARY; PRT; 201 AA.

AC O9RJZ8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE PUTATIVE AMIDASE.  
 GN SCG11A.03.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinash H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL133210; CAB61584.1; -.  
 DR InterPro: IPR002502; Amidase\_2.  
 DR Pfam: PF01510; Amidase\_2; 1.  
 SQ SEQUENCE 201 AA; 22749 MW; B8EF477E06A20468 CRC64;

Query Match 87.2%; Score 34; DB 2; Length 201;  
 Best Local Similarity 85.7%; Pred. No. 14;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 DB 176 EYVGSQH 182

## RESULT 11

O9NP06 PRELIMINARY; PRT; 192 AA.

AC O9NP06;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
 DE LIP, LCAT-LIKE LYSOPHOSPHOLIPASE (FRAGMENT).  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Puvion R., Estivill X., Escarceller M., Sumoy L.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.

RA Aufray C., Ansoorge W., Ballabio A., Estivill X., Gibson K.,  
 RA Lehnach H., Poustka A., Lundeberg J.;  
 RT "The European IMGE consortium for integrated Molecular analysis of  
 human gene transcripts.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL389577; CAB97531.1; -.  
 FT NON\_TER  
 SQ SEQUENCE 192 AA; 21609 MW; 04A7AE8CB344F213 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 192;  
 Best Local Similarity 71.4%; Pred. No. 21;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 DB 166 EYVGSQH 172

## RESULT 12

O9UG04 PRELIMINARY; PRT; 272 AA.

AC O9UG04;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE HYPOTHEICAL 31.0 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC.12657).  
 GN DKF2P564A0122.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Wamutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN, NEUROBLASTOMA;  
 RA Strusberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL110209; CAB53675.1; -.  
 DR EMBL: BC011640; AAH11640.1; -.  
 DR InterPro: IPR003386; LACT.  
 DR Pfam: PF02450; LACT; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 272 AA; 31016 MW; ACCC5E1680D7A720 CRC64;

Query Match 84.6%; Score 33; DB 4; Length 272;  
 Best Local Similarity 71.4%; Pred. No. 31;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 DB 246 EYVGSQH 252

## RESULT 13

O80418 PRELIMINARY; PRT; 367 AA.

AC O80418;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

```

DE NTC16 PROTEIN.
GN NTC16.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka-Ueguchi M., Itoh H., Oyama N., Koshioka M., Matsuo M.;
RT "Over-expression of a tobacco homeobox gene, NTH15, decreases the
RT expression of a gibberellin biosynthetic gene encoding GA 20-
RT oxidase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB016084; BAA31690.1; -;
DR InterPro: IPR002419; Fe_Asc_oxidored; 1.
DR Pfam: PF00671; Fe_Asc_oxidored; 1.
SQ SEQUENCE 367 AA; 42170 MW; 923BC90B3BBAC05 CRC64;

Query Match      84.6%; Score 33; DB 10; Length 367;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
DB 149 EVPGSOH 155

RESULT 14
QY2B3
ID QY2B3 PRELIMINARY; PRT; 412 AA.
AC QY2B3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LCAT-LIKE LYSOPHOSPHOLIPASE (LLPL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99194552; PubMed=10092508;
RA Tanigawa Y., Shibata S., Kita S., Horikoshi K., Shirafuji H.,
RA Sumino Y., Fujino M.;
RT "Cloning and expression of a novel lysophospholipase which
RT structurally resembles lecithin cholesterol acyltransferase."
RL Biochem. Biophys. Res. Commun. 257:50-56(1999).
DR EMBL: AB017494; BAA76877.1; -;
DR InterPro: IPR003386; LACT.
DR Pfam: PF02450; LACT; 1.
SQ SEQUENCE 412 AA; 46657 MW; 1FEBA8A5783AF050A CRC64;

Query Match      84.6%; Score 33; DB 4; Length 412;
Best Local Similarity 71.4%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
DB 386 ELPGSEH 392

RESULT 15
QYVE2
ID QYVE2 PRELIMINARY; PRT; 868 AA.
AC QYVE2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG3837 PROTEIN.
GN CG3837.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durlin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Stokrop R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2185(2000).
DR EMBL: AE003706; AAE5118.1; -;
DR FlyBase: FBgn0038279; CG3837.
DR InterPro: IPR000494; EGFR_L.
DR Pfam: PF01030; Recep_L_domain; 2.
SQ SEQUENCE 868 AA; 98349 MW; F6562A64E72E7B21 CRC64;

Query Match      84.6%; Score 33; DB 5; Length 868;
Best Local Similarity 71.4%; Pred. No. 116+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSOH 7
DB 812 ELPGRQH 818

Search completed: October 24, 2002, 15:22:23
Job time : 12.973 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 16:36:21 ; Search time 19.88 Seconds  
(without alignments)  
12.062 Million cell updates/sec

Title: US-09-786-648-2

Perfect score: 39

Sequence: 1 EVPGSQH 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt-39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	100.0	124	1 CHTB_VIBCH	P01556 vibrio chol
2	39	100.0	124	1 ELBH_ECOLI	P13811 escherichia
3	39	100.0	124	1 ELBP_ECOLI	P32890 escherichia
4	32	82.1	455	1 GUDH_BACSU	P42238 bacillus su
5	32	82.1	534	1 YOGI_CAEEL	P34610 caenorhabdi
6	32	82.1	666	1 PD14_MOUSE	O92183 mus musculu
7	32	82.1	765	1 SIM1_MOUSE	O61045 mus musculu
8	31	79.5	137	1 Y1SK_MYCLE	P13733 mycobacteri
9	31	79.5	219	1 CIDA_HUMAN	O60543 homo sapien
10	31	79.5	273	1 DLX3_NOTVI	P53770 notophthalm
11	31	79.5	274	1 DLX3_PLEMA	O91284 pleurodeles
12	31	79.5	280	1 DLX3_AMBME	O90229 ambystoma m
13	31	79.5	414	1 COR2_BOVIN	O91C17 bos taurus
14	31	79.5	414	1 COR2_HUMAN	P24468 homo sapien
15	31	79.5	414	1 COR2_MOUSE	P43135 mus musculu
16	31	79.5	414	1 COR2_RAT	O09018 rattus norv
17	31	79.5	432	1 AROC_NEUCR	O12640 neuropeptid
18	31	79.5	459	1 IL7R_MOUSE	P16872 mus musculu
19	31	79.5	459	1 IL7R_HUMAN	P16871 homo sapien
20	30	76.9	182	1 KPTA_PSEAE	P77989 thermoaer
21	30	76.9	336	1 RPOA_PSEAE	O91778 pseudomonas
22	30	76.9	376	1 AROC_YEAST	P28777 saccharomyc
23	30	76.9	459	1 IL7R_HUMAN	P16871 homo sapien
24	30	76.9	500	1 GABR_HUMAN	P80404 homo sapien
25	30	76.9	504	1 A37C_DROME	P18487 drosophila
26	30	76.9	573	1 AMH2_HUMAN	O16671 homo sapien
27	30	76.9	616	1 RFX5_HUMAN	P48382 homo sapien
28	30	76.9	622	1 SR68_CAEEL	O20822 caenorhabdi
29	30	76.9	766	1 SIM1_HUMAN	P81133 homo sapien
30	30	76.9	775	1 TH11_SCHPO	P36598 schizosacch
31	30	76.9	805	1 AHR_MOUSE	P30561 mus musculu
32	30	76.9	808	1 PLD_TOBAC	P93400 nicotiana t
33	30	76.9	853	1 AHR_RAT	P41738 rattus norv

34	30	76.9	855	1 ST14_HUMAN	O9556 homo sapien
35	30	76.9	954	1 DRP2_HUMAN	O13474 homo sapien
36	30	76.9	1286	1 PATC_DROME	P18502 drosophila
37	30	76.9	1538	1 LHR_ECOLI	P30015 escherichia
38	30	76.9	1638	1 BRM_DROME	P25439 drosophila
39	30	76.9	1808	1 TENK_CHICK	P10039 gallus gall
40	29	74.4	277	1 ZDKG_CORSP	P15339 corynebacte
41	29	74.4	395	1 INX3_DROME	O9vas7 drosophila
42	29	74.4	404	1 Y4XM_RHISN	P55705 rhizobium s
43	29	74.4	586	1 RRP0_BMYVF	P09507 beet weater
44	29	74.4	621	1 HEMI_AGABI	O92403 agaricus bi
45	29	74.4	773	1 MAK5_YEAST	P38112 saccharomyc

## ALIGNMENTS

RESULT ID	CTHB_VIBCH	STANDARD	PRT	124 AA.
AC	P01556; Q9J002;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	CHOLERA ENTEROTOXIN, BETA CHAIN PRECURSOR.			
GN	CTXB OR TOXB OR VC1456.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=84061784; PubMed=6315707;			
RA	Lockman H., Kaper J.B.;			
RT	"Nucleotide sequence analysis of the A2 and B subunits of Vibrio cholerae enterotoxin.";			
RL	J. Biol. Chem. 258:13722-13726(1983).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=EL TOR 2125;			
RA	MEDLINE=84068199; PubMed=6646234;			
RT	Mekalanos J.J., Swartz D.J., Pearson G.D.N., Harford N., Groyne F., de Wilde M.;			
RL	"Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development.";			
RN	Nature 306:551-557(1983).			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=EL TOR 2125;			
RA	Dams E., de Wolf M., Dierick W.;			
RT	Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=4260B / SEROTYPE O139;			
RX	MEDLINE=94237453; PubMed=8181723;			
RA	Lebens M., Holmgren J.;			
RT	"Structure and arrangement of the cholera toxin genes in Vibrio cholerae O139.";			
RL	PEMS Microbiol. Lett. 117:197-202(1994).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=1854 / O139-BENGAL;			
RA	Yamamoto K., Do V.G.F., Xu M., Iida T., Miwatani T., Albert M.J., Honda T.;			
RT	Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=EL TOR N16961 / SEROTYPE O1;			
RX	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,			

RA Salzbberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 RN [7]  
 RX MEDLINE-7800537; PubMed-903363;  
 RX Kurosky A., Markel D.E., Peterson J.W.;  
 RT "Covalent structure of the beta chain of cholera enterotoxin.";  
 RL J. Biol. Chem. 252:7257-7264(1977).  
 RN [8]  
 RX MEDLINE-7800536; PubMed-903362;  
 RX Lai C.-Y.;  
 RT "Determination of the primary structure of cholera toxin B subunit.";  
 RL J. Biol. Chem. 252:7249-7256(1977).  
 RN [9]  
 RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE-94272319; PubMed-8003954;  
 RA Merritt E.A., Sarfaty S., van den Akker F., L'Holr C., Martial J.A.,  
 RA Hol W.G.J.;  
 RT "Crystal structure of cholera toxin B-pentamer bound to receptor GM1  
 RT pentasaccharide.";  
 RL Protein Sci. 3:166-175(1994).  
 RN [10]  
 RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX MEDLINE-95387994; PubMed-7658472;  
 RA Zhang R.G., Westbrook M.L., Westbrook E.M., Scott D.L., Orlowski Z.,  
 RA Maulik P.R., Reed R.A., Shipley G.G.;  
 RT "The 2.4 A crystal structure of cholera toxin B subunit pentamer:  
 RT cholegeroid.";  
 RL J. Mol. Biol. 251:550-562(1995).  
 RN [11]  
 RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX STRAIN-OGAWA 41 / CLASSICAL BIOTYPE;  
 RX MEDLINE-97376625; PubMed-9232653;  
 RA Merritt E.A., Sarfaty S., Jobling M.G., Chang T., Holmes R.K.,  
 RA Hirst T.R., Hol W.G.J.;  
 RT "Structural studies of receptor binding by cholera toxin mutants.";  
 RL Protein Sci. 6:1516-1528(1997).  
 CC -1- FUNCTION: THE BETA CHAIN AGGREGATE (B SUBUNIT) IS INVOLVED IN  
 CC BINDING TO CELL MEMBRANES.  
 CC -1- SUBUNIT: CONTAINS 3 KINDS OF CHAINS. AN ALPHA AND A GAMMA CHAIN  
 CC (FROM THE SAME PRECURSOR MOLECULE), LINKED BY AN INTERCHAIN  
 CC DISULFIDE BOND, ASSOCIATE NONCOVALENTLY WITH AN AGGREGATE OF 4 TO  
 CC 6 BETA CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isp-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; X01171; CAA24996.1; -;  
 DR EMBL; K01170; AAA27573.1; -;  
 DR EMBL; D30053; BAA06291.1; -;  
 DR EMBL; X58786; CAA41593.1; -;  
 DR EMBL; X76390; CAA53873.1; -;  
 DR EMBL; X76391; CAA53976.1; -;  
 DR EMBL; AE004224; AAF94613.1; -;  
 DR PIR; A01819; XVCB;  
 DR PIR; A05130; A05130;  
 DR PIR; S14624; S14624;  
 DR PDB; 2CHB; 03-DEC-97.  
 DR PDB; 3CHB; 12-AUG-98.  
 DR PDB; 1CHP; 08-MAR-96.  
 DR PDB; 1CHQ; 08-MAR-96.  
 DR PDB; 1FGB; 23-DEC-96.  
 DR PDB; 1XPB; 01-APR-97.  
 DR PDB; 1XTC; 01-AUG-96.

DR PDB; 1CT1; 15-OCT-97.  
 DR TIGR; VC1456; -;  
 DR InterPro; IPR001835; -;  
 DR Pfam; PF01376; Enterotoxin\_B; 1.  
 DR PRINTS; PR00772; ENTEROTOXINB.  
 KW Membrane; Enterotoxin; Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 1 21  
 FT DISULFID 30 107  
 FT CONFLICT 33 33  
 FT CONFLICT 39 39  
 FT CONFLICT 43 43  
 FT CONFLICT 68 68  
 FT CONFLICT 70 70  
 FT CONFLICT 75 75  
 FT CONFLICT 91 91  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT TURN 34 35  
 FT TURN 36 44  
 FT STRAND 47 51  
 FT TURN 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT HELIX 80 99  
 FT TURN 100 100  
 FT STRAND 102 109  
 FT STRAND 115 123  
 SQ SEQUENCE 124 AA; 13957 MW; 9AA393E3EA8E3EBF CRC64;  
 Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EYVGSOH 7  
 Db 72 EYVGSOH 78  
 |||||  
 RESULT 2  
 ELBH\_ECOLI  
 ID ELBH\_ECOLI STANDARD; PRT; 124 AA.  
 AC P13811;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B, HUMAN) (LTB-B).  
 GN EYV OR LTPB.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-ISOLATE H74-114;  
 RC MEDLINE=85156481; PubMed=3884513;  
 RX Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT cistrons from Escherichia coli of human and porcine origin.";  
 RL Infect. Immun. 48:73-77(1985).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-ISOLATE H10407;  
 RC MEDLINE=83114628; PubMed=6759877;  
 RX Yamamoto T., Tamura T.A., Yokota T., Takano T.;  
 RA "Overlapping genes in the heat-labile enterotoxin operon originating  
 RT from Escherichia coli human strain.";  
 RL Mol. Gen. Genet. 188:356-359(1982).  
 RN [3]  
 RN SEQUENCE FROM N.A.

RC STRAIN-ISOLATE H10407;  
 RA MEDLINE-93252225; PubMed-8486242;  
 RA Inoue T., Tsuji T., Koto M., Imamura S., Miyama A.;  
 RT "Amino acid sequence of heat-labile enterotoxin from chicken  
 RT enterotoxigenic Escherichia coli is identical to that of human strain  
 RT H 10407.";  
 RL FEMS Microbiol. Lett. 108:157-161(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ETEC LT 87;  
 RA Germani Y., Desperrier J.M.;  
 RN Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE-95349400; PubMed-7623669;  
 RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappunli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.04 ANGSTROMS).  
 RX MEDLINE-99185101; PubMed-10085117;  
 RA Matkovic-Galogovic D., Loregian A., D'Acunzio M.R., Battistutta R.,  
 RA Tossi A., Pail G., Zanotti G.;  
 RT "Crystal structure of the B subunit of escherichia coli heat-labile  
 RT enterotoxin carrying peptides with anti-herpes simplex virus type 1  
 RT activity.";  
 RL J. Biol. Chem. 274:8764-8769(1999).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M1874; AAA98064.1; -;  
 DR EMBL: J01646; AAB02982.1; -;  
 DR EMBL: S60731; AAC60441.1; -;  
 DR EMBL: X83966; CAA58800.1; -;  
 DR PDB: 1LFR; 23-MAR-99.  
 DR InterPro: IPR001835; -;  
 DR Pfam: PF01376; Enterotoxin\_B; 1.  
 DR PRINTS: PRO0772; ENTEROTOXINB.  
 DR Enterotoxin; Signal: 3D-structure.  
 KW SIGNAL 1 21  
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT VARIANT 6 6 F -> C (IN ISOLATE H10407).  
 FT VARIANT 20 20 H -> Y (IN ISOLATE H10407).  
 FT VARIANT 34 34 H -> R (IN ISOLATE H10407).  
 SQ SEQUENCE 124 AA; 14027 MW; E9FF7C7B9D3BC47 CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYVPSOH 7  
 DB 72 EYVPSOH 78

RESULT 3  
 ID EMBL ECOLI STANDARD; PRT; 124 AA.  
 AC P32830; P13768; P01557;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HEAT-LABILE ENTEROTOXIN B CHAIN PRECURSOR (LT-B, PORCINE) (LTP-B).  
 GN ELTB OR LTPB.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE P307;  
 RX MEDLINE-81074965; PubMed-7003397;  
 RA Dallas W.S., Falkow S.;  
 RT "Amino acid sequence homology between cholera toxin and Escherichia  
 RT coli heat-labile toxin.";  
 RL Nature 288:499-501(1980).  
 RN [2]  
 RP REVISIONS TO 28 AND 64.  
 RC STRAIN-ISOLATE P307;  
 RX MEDLINE-85156481; PubMed-3884513;  
 RA Leong J., Vinal A.C., Dallas W.S.;  
 RT "Nucleotide sequence comparison between heat-labile toxin B-subunit  
 RT cistrons from Escherichia coli of human and porcine origin.";  
 RL Infect. Immun. 48:73-77(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE PCG86;  
 RX MEDLINE-87137303; PubMed-3546273;  
 RA Yamamoto T., Gojohori T., Yokota T.;  
 RT "Evolutionary origin of pathogenic determinants in enterotoxigenic  
 RT Escherichia coli and Vibrio cholerae O1.";  
 RL J. Bacteriol. 169:1352-1357(1987).  
 RN [4]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RX MEDLINE-87280041; PubMed-3301830;  
 RA Ibrahim I., Gentz R.;  
 RT "A functional interaction between the signal peptide and the  
 RT translation apparatus is detected by the use of a single point  
 RT mutation which blocks translocation across mammalian endoplasmic  
 RT reticulum.";  
 RL J. Biol. Chem. 262:10189-10194(1987).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE-93240541; PubMed-8478941;  
 RA Sixma T.K., van Zanten B.A.M., Dauber Z., Hol W.G.J.;  
 RT "Refined structure of Escherichia coli heat-labile enterotoxin, a  
 RT close relative of cholera toxin.";  
 RL J. Mol. Biol. 230:890-918(1993).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE-91238966; PubMed-2034287;  
 RA Sixma T.K., Pronk S.E., Kalk K.H., Wartna E.S., van Zanten B.A.M.,  
 RA Witboldt B., Hol W.G.J.;  
 RT "Crystal structure of a cholera toxin-related heat-labile enterotoxin  
 RT from E. coli.";  
 RL Nature 351:371-377(1991).  
 RN [7]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE-95349400; PubMed-7623669;  
 RA Domenighini M., Pizsa M., Jobling M.G., Holmes R.K., Rappunli R.;  
 RT "Identification of errors among database sequence entries and  
 RT comparison of correct amino acid sequences for the heat-labile  
 RT enterotoxins of Escherichia coli and Vibrio cholerae.";  
 RL Mol. Microbiol. 15:1165-1167(1995).  
 CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY  
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.  
 CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: M17873; AAA98065.1; -  
 DR EMBL: M15363; AAA24792.1; -  
 DR EMBL: M17101; AAA23973.1; -  
 DR PIR: A01820; QLECEB.  
 DR PIR: B26946; QLECEB.  
 DR PDB: 1LTA; 31-JAN-94.  
 DR PDB: 1LTB; 31-JAN-94.  
 DR PDB: 1LTG; 15-SEP-95.  
 DR PDB: 1LTI; 17-AUG-96.  
 DR PDB: 1LTS; 31-JAN-94.  
 DR PDB: 1LTT; 31-JAN-94.  
 DR PDB: 1LTK; 07-JUL-97.  
 DR PDB: 1LTA; 16-JUN-97.  
 DR PDB: 1LTS; 16-DEC-97.  
 DR PDB: 1LTK; 03-DEC-97.  
 DR PDB: 1LTI; 20-APR-95.  
 DR InterPro: IPR001835; -  
 DR Pfam: PF01376; Enterotoxin\_B; 1.  
 DR PRINTS: PR00772; ENTEROTOXINB.  
 KW Enterotoxin; Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 124 HEAT-LABILE ENTEROTOXIN B CHAIN.  
 FT DISULFID 30 107  
 FT HELIX 26 30  
 FT TURN 31 32  
 FT STRAND 36 43  
 FT STRAND 47 51  
 FT TURN 54 55  
 FT STRAND 58 62  
 FT TURN 64 65  
 FT STRAND 68 71  
 FT TURN 76 77  
 FT HELIX 80 98  
 FT TURN 99 100  
 FT STRAND 103 109  
 FT STRAND 115 123  
 SQ SEQUENCE 124 AA; 14133 MW; 6DB7DE58395EA70D CRC64;

Query Match 100.0%; Score 39; DB 1; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 DB 72 EYVGSQH 78

RESULT 4  
 GUDH\_BACSU STANDARD; PRT; 455 AA.  
 AC P42238; -  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE GLUCARATE DEHYDRATASE (EC 4.2.1.40) (GDH) (GLUCD).  
 GN GUDH.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95219079; PubMed=7704254;  
 RA Ogawa K., Akagawa E., Nakamura K., Yamane K.;  
 RT "Determination of a 21548 bp nucleotide sequence around the 24  
 RT degrees region of the Bacillus subtilis chromosome."  
 RL Microbiology 141:269-275(1995).  
 CC -1- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-

DEOXY-D-GLUCARATE (5-KDGLUC) (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-GLUCARATE = 5-DEHYDRO-4-DEOXY-D-GLUCARATE + H(2)O.  
 CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.  
 CC -1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: D30808; BAA06470.1; -  
 DR EMBL: 299105; CAB12043.1; -  
 DR Subtilist; BG11161; guld.  
 KW Lyase.  
 SQ SEQUENCE 455 AA; 50782 MW; 3238486007698C2A CRC64;

Query Match 82.1%; Score 32; DB 1; Length 455;  
 Best Local Similarity 71.4%; Pred. No. 27;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7  
 DB 64 EYVGSQH 70

RESULT 5  
 YOGI\_CAEEL STANDARD; PRT; 534 AA.  
 ID YOGI\_CAEEL  
 AC P34610;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE HYPOTHETICAL 59.2 KDA PROTEIN ZK112.1 IN CHROMOSOME III.  
 GN ZK112.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=62339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rikken L., Koopra A., Saunders D., Showman R.,  
 RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,  
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: L14324; AAA28181.1; -  
 DR PIR: S44886; S44886.



DR WormPep; ZK112.1; CE00372.  
 DR InterPro; IPR000073; -  
 DR Pfam; PF00561; abhydrolase; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 534 AA; 59221 MW; C43F0104B42E4630 CRC64;

Query Match 82.1%; Score 32; DB 1; Length 534;  
 Best Local Similarity 71.4%; Pred. No. 32;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 1:1111  
 DB 442 EIPGSAH 448

RESULT 6  
 PD14\_MOUSE STANDARD; PRT; 666 AA.  
 ID PD14\_MOUSE Q92183;  
 AC Q92183;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROTEIN-ARGININE DEIMINASE TYPE IV (EC 3.5.3.15) (PEPTIDYLARGININE  
 DE DEIMINASE IV).  
 GN PD14.  
 OS Mus musculus (Mouse).  
 GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Epidermis;  
 RX MEDLINE=99192810; PubMed=10092850;  
 RA Rusd A.A., Ikejiri Y., Ono H., Yonekawa T., Shiraiwa M., Kawada A.,  
 Takahara H.;  
 RT "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type  
 RT I, type III and type IV, and the expression pattern of type I in  
 RT mouse.";  
 RL Eur. J. Biochem. 259:660-669(1999).  
 CC -1- FUNCTION: CATALYZES THE DEIMINATION OF ARGININE RESIDUES OF  
 CC PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN L-ARGININE + H(2)O = PROTEIN  
 CC L-CITRULLINE + NH(3).  
 CC -1- COFACTOR: REQUIRES CALCIUM IONS.  
 CC -1- SIMILARITY: SOME SIMILARITY WITH ARGINASE AND TO ARGININE  
 CC DEIMINASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB013850; BAA34246.1; -  
 DR MGD: MGI:1338898; Pd14.  
 KW Hydrolase; Calcium-binding; Multigene family.  
 FT CA BIND 505 516 POTENTIAL.  
 FT SEQUENCE 666 AA; 74476 MW; 70FAE4E7E332D34A CRC64;

Query Match 82.1%; Score 32; DB 1; Length 666;  
 Best Local Similarity 71.4%; Pred. No. 40;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 1:1111  
 DB 239 EIPGSOH 245

RESULT 7

SIM1\_MOUSE  
 ID SIM1\_MOUSE STANDARD; PRT; 765 AA.  
 AC 061045; P70183;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE SINGLE-MINDED HOMOLOG 1 (MSIM1).  
 GN SIM1.  
 OS Mus musculus (Mouse).  
 GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SWISS WEBSTER;  
 RX MEDLINE=97020303; PubMed=8812055;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RT "Expression patterns of two murine homologs of Drosophila  
 RT single-minded suggest possible roles in embryonic patterning and in  
 RT the pathogenesis of Down syndrome.";  
 RL Mol. Cell. Neurosci. 7:1-16(1996).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=97029422; PubMed=8875433;  
 RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,  
 RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,  
 RA Tessier-Lavigne M.;  
 RL Mol. Cell. Neurosci. 7:519-519(1996).  
 RN [3]  
 RP REVISIONS TO C-TERMINUS.  
 RX MEDLINE=97343329; PubMed=919934;  
 RA Fan C.-M.;  
 RL Unpublished results, cited by:  
 RL Charast R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,  
 RL Wang Y., Shmidt N., Antonarakis S.E.;  
 RL Genome Res. 7:615-624(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J;  
 RX MEDLINE=96413359; PubMed=8927054;  
 RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,  
 RA Saitoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;  
 RT "Two new members of the murine sim gene family are transcriptional  
 RT repressors and show different expression patterns during mouse  
 RT embryogenesis.";  
 RL Mol. Cell. Biol. 16:5865-5875(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RA Hosoya T.;  
 RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SUBUNIT.  
 RX MEDLINE=97172525; PubMed=9020169;  
 RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;  
 RT "Two murine homologs of the Drosophila single-minded protein that  
 RT interact with the mouse aryl hydrocarbon receptor nuclear  
 RT translocator protein.";  
 RL J. Biol. Chem. 272:4451-4457(1997).  
 CC -1- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS  
 CC DURING EMBRYOGENESIS AND IN THE ADULT.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.  
 CC -1- TISSUE SPECIFICITY: DETECTED IN LUNG, SKELETAL MUSCLE AND KIDNEY.  
 CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING  
 CC KIDNEY, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING  
 CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U40575; AAA91201.1; ALT\_SEQ.  
DR EMBL; D79209; BAA11467.1; -.  
DR EMBL; AB013491; BAA28270.1; -.  
DR EMBL; AB013484; BAA28270.1; JOINED.  
DR EMBL; AB013485; BAA28270.1; JOINED.  
DR EMBL; AB013486; BAA28270.1; JOINED.  
DR EMBL; AB013487; BAA28270.1; JOINED.  
DR EMBL; AB013488; BAA28270.1; JOINED.  
DR EMBL; AB013489; BAA28270.1; JOINED.  
DR EMBL; AB013490; BAA28270.1; JOINED.  
DR MGI; MGI:98306; Sim1.  
DR InterPro; IPR000014; -.  
DR InterPro; IPR001067; -.  
DR InterPro; IPR001610; -.  
DR InterPro; IPR003015; -.  
DR Pfam; PF00785; PAC; 1.  
DR Pfam; PF00989; PAS; 2.  
DR PRINTS; PR00785; NCTRNSLOCATR.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
KW Transcription regulation; DNA-binding.  
FT DNA\_BIND 1 13 BASIC DOMAIN.  
FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT REPEAT 98 148 PAS-1.  
FT REPEAT 239 289 PAS-2.  
FT CONFLICT 133 133 H->L (IN REF. 1).  
FT CONFLICT 176 176 MISSING (IN REF. 1).  
FT CONFLICT 322 322 P->R (IN REF. 1).  
FT CONFLICT 480 480 A->P (IN REF. 1).  
FT CONFLICT 537 537 D->S (IN REF. 1).  
SQ SEQUENCE 765 AA; 85540 MW; B1A7F7DA857BCD17 CRC64;

Query Match Best Local Similarity 82.1%; Score 32; DB 1; Length 765;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYVGSQH 7  
DB 423 ERPGSQH 429

RESULT 8  
ID Y15K\_MYCLE STANDARD: PRT; 137 AA.  
Y15K\_MYCLE  
AC P13733;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHEETICAL 15 KDA PROTEIN IN 18 KDA ANTIGEN 3'REGION.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1169;  
RN (1)  
RP MEDLINE=88088878; PubMed=2447183;  
RA Booth R.J., Harris D.P., Love J.M., Watson J.D.;  
RT "Antigenic proteins of Mycobacterium leprae. Complete sequence of the  
RT gene for the 18-kDa protein";  
RL J. Immunol. 140:597-601(1988).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M19058; AAA88230.1; -.  
DR PIR; B27586; B27586.  
KW Hypothetical protein.  
SQ SEQUENCE 137 AA; 15195 MW; 633E68BF3FDAD0B CRC64;

Query Match Best Local Similarity 79.5%; Score 31; DB 1; Length 137;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
DB 127 VPGSQH 132

RESULT 9  
ID CIDA\_HUMAN STANDARD: PRT; 219 AA.  
CIDA\_HUMAN  
AC 060543;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELL DEATH ACTIVATOR CIDE-A (CELL DEATH-INDUCING DEFA-LIKE EFFECTOR  
DE A).  
GN CIDEA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP MEDLINE=98232498; PubMed=9564035;  
RA Inohara N., Koseki T., Chen S., Wu X., Nunez G.;  
RT "CIDE, a novel family of cell death activators with homology to the 45  
RT kDa subunit of the DNA fragmentation factor";  
RL EMO J. 17:2526-2533(1998).  
CC -1- FUNCTION: ACTIVATES APOPTOSIS.  
CC -1- SUBUNIT: INHIBITED BY DFFB.  
CC -1- SIMILARITY: CONTAINS 1 CIDE-N DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF041378; AAC34987.1; -.  
DR MIM; 604460; -.  
DR PIR; P02017; CIDE-N; 1.  
KW Apoptosis.  
FT DOMAIN 33 110 CIDE-N.  
SQ SEQUENCE 219 AA; 24686 MW; 05F704823CE71C0E CRC64;

Query Match Best Local Similarity 79.5%; Score 31; DB 1; Length 219;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPGSQH 7  
DB 109 MPGSQH 114

RESULT 10  
ID DLX3\_NOTVT STANDARD: PRT; 273 AA.  
DLX3\_NOTVT

```

AC P53770;
DR 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
GN DLX3 OR BOX4.
OS Notopterus viridescens (Eastern newt) (Triturus viridescens);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;
OC Notopterus.
NCBI_TaxID=8316;
RX MEDLINE=9350784; PubMed=1358728;
RA Beauchemin M., Savard P.;
RT "Two distal-less related homeobox-containing genes expressed in
RL regeneration blastemas of the newt.";
CC Dev. Biol. 134:55-65(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN REGENERATING BLASTEMAS. DETECTED
CC IN FORELIMBS, HINDLIMBS, THE TAIL, FLANK, AND BRAIN AS WELL AS IN
CC LIMB AND TAIL BLASTEMAS.
CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63531; CAA45094.1; -
DR HSP: P02836; 1HDD.
DR InterPro: IPR000047; -
DR InterPro: IPR001356; -
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTREPRESSR.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 125 184
SQ SEQUENCE 273 AA; 30654 MW; B356D01233061F2F CRC64;

Query Match 79.5%; Score 31; DB 1; Length 273;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPSGSH 7
DB 188 EYPSGSH 194

RESULT 11
DLX3_PLEWA STANDARD; PRT; 274 AA.
ID DLX3_PLEWA
AC 091284;
DR 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
GN DLX3.
OS Pleurodeles waltl (Iberian ribbed newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae;
OC Pleurodeles.
NCBI_TaxID=8319;
RX [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=96391186; PubMed=8798159;
RA Nicolas S., Massacrier A., Caubit X., Cau P., le Parco Y.;
RT "A distal-less-like gene is induced in the regenerating central
RL nervous system of the urodele pleurodeles waltl.";
CC Mech. Dev. 56:209-220(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE REGENERATING SPINAL CORD BUT
CC NOT IN THE ADULT ONE.
CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49645; AAA98645.1; -
DR HSP: P02836; 1HDD.
DR InterPro: IPR000047; -
DR InterPro: IPR001356; -
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTREPRESSR.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 126 185
SQ SEQUENCE 274 AA; 30607 MW; 9418553BA80B8C8 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 274;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPSGSH 7
DB 189 EYPSGSH 195

RESULT 12
DLX3_AMBME STANDARD; PRT; 280 AA.
ID DLX3_AMBME
AC 090229;
DR 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN HOMEBOX PROTEIN DLX-3.
OS Ambystoma mexicanum (Axolotl).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Ambystomatidae;
OC Ambystoma.
NCBI_TaxID=8296;
RX [1]
RP SEQUENCE FROM N.A.
DE MEDLINE=97108743; PubMed=8951064;
RA Mullen L.M., Bryant S.V., Torok M.A., Blumberg B., Gardiner D.M.;
RT "Nerve dependency of regeneration: the role of Distal-less and FGF
RL signaling in amphibian limb regeneration.";
CC Dev. Biol. 122:3487-3497(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: U59480: AAB49668.1; -.
CC HSSP: P02836: 1HDD.
CC InterPro: IPR000047; -.
CC InterPro: IPR001356; -.
CC Pfam: PF000046: homeobox.1.
CC PRINTS: PR000024: HOMEBOX.
CC PRINTS: PR000031: HTHREPRESSOR.
CC PROSITE: PS00027: HOMEBOX_1; 1.
CC PROSITE: PS00071: HOMEBOX_2; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
KW DNA_BIND
FT DNA_BIND 126 185 HOMEBOX.
SQ SEQUENCE 280 AA; 31200 MW; 0F8183097ABAC791 CRC64;

Query Match
Best Local Similarity 79.5%; Score 31; DB 1; Length 280;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYVGSQH 7
Db 189 EYVGMH 195

RESULT 13
COT2_BOVIN
ID COT2_BOVIN STANDARD; PRT; 414 AA.
AC 09TMR7.
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COUP TRANSCRIPTION FACTOR 2 (COUP-TF2) (COUP-TF II).
GN NR2P2 OR TFCOUP2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Corpus luteum;
RA Walther N.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
CC BINDS TO DNA SITE A (BY SIMILARITY).
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ249441: CAB55624.1; -.
CC InterPro: IPR000536; -.
CC InterPro: IPR001628; -.
CC InterPro: IPR001723; -.
CC InterPro: IPR003068; -.
CC Pfam: PF00104: hormone_rec.1.
CC Pfam: PF00105: zf-C4.1.
CC PRINTS: PR00047: STROIPFINGER.
CC PRINTS: PR00398: STROHORMER.
CC PRINTS: PR01282: COUPTNFACTOR.
CC PROSITE: PS00031: NUCLEAR RECEPTOR; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT DOMAIN 71 75 POLY-GLN.

```

```

FT DNA_BIND 79 144 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 79 99 C4-TYPE.
FT ZN_FING 115 139 C4-TYPE.
FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION (BY
FT SIMILARITY).
SQ SEQUENCE 414 AA; 45557 MW; C24CB023C8A27F57 CRC64;

Query Match
Best Local Similarity 79.5%; Score 31; DB 1; Length 414;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYVGSQ 6
Db 14 EYVGSQ 19

RESULT 14
COT2_HUMAN
ID COT2_HUMAN STANDARD; PRT; 414 AA.
AC P24468: Q03754;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COUP TRANSCRIPTION FACTOR 2 (COUP-TF2) (COUP-TF II) (APOLIPOPROTEIN AI
DE REGULATORY PROTEIN-1) (ARP-1).
GN NR2P2 OR TFCOUP2 OR ARP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91118002; PubMed=1899293;
RA Ladas J.A.A., Karathanasis S.K.;
RT "Regulation of the apolipoprotein AI gene by ARP-1, a novel member of
RT the steroid receptor superfamily."
RL Science 251:561-565(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Speckmayer R.W.M., Paulweber B., Sandhofer F.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-351 FROM N.A.
RX MEDLINE=92314709; PubMed=1820218;
RA Wang L.H., Ing N.H., Tsai S.Y., O'Malley B.W., Tsai M.J.;
RT "The COUP-TFs compose a family of functionally related transcription
RT factors."
RL Gene Expr. 1:207-216(1991).
CC -1- FUNCTION: REGULATION OF THE APOLIPOPROTEIN AI GENE TRANSCRIPTION.
CC BINDS TO DNA SITE A.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M64497: AAB86429.1; -.
CC EMBL: U60477: AAB09475.1; -.
CC EMBL: M62760: AAB21479.1; -.
CC PIR: A37133; A37133.
CC HSSP: P19793: 2NLL.
CC TRANSPAC: T00045; -.
CC MIM: 107773; -.
CC InterPro: IPR000536; -.

```

DR InterPro: IPR001628; -  
 DR InterPro: IPR001723; -  
 DR InterPro: IPR003068; -  
 DR Pfam: PF00104; hormone\_rec; 1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR PRINTS: PR00398; STRDHOMONER.  
 DR PRINTS: PR01282; COUPINFACOR.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Transcription regulation; DNA-binding; Nuclear protein;  
 KM Zinc-finger.  
 FT DOMAIN 71 75 POLY-GIN.  
 FT DNA\_BIND 79 144 NUCLEAR\_RECEPTOR-TYPE.  
 FT ZN\_FING 79 99 C4-TYPE.  
 FT ZN\_FING 115 139 C4-TYPE.  
 FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION.  
 SQ SEQUENCE 414 AA; 45571 MW; C24CB2EBC8A27EBC CRC64;

Query Match 79.5%; Score 31; DB 1; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
 |||||  
 DB 14 EVPGSQ 19

RESULT 15  
 COT2\_MOUSE STANDARD; PRT; 414 AA.  
 ID COT2\_MOUSE  
 AC P43135;  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE COUP TRANSCRIPTION FACTOR 2 (COUP-TF2) (COUP-TF II) (APOLIPROTEIN AI  
 DE REGULATORY PROTEIN-1) (ARP-1).  
 GN NR2F2 OR TFCOUP2 OR ARP1 OR ARP-1 OR APORP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95034311; PubMed=7947324;  
 RA Jonk L.J.C., de Jonge M.J., Pals C.E.G.M., Missink S.,  
 RA Vermaat J.M.A., Schoorlemmer J., Kruijer W.,  
 RT Cloning and expression during development of three murine members of  
 RT the COUP family of nuclear orphan receptors.";  
 RL Mech. Dev. 47:81-97(1994).  
 CC -1- FUNCTION: REGULATION OF THE APOLIPROTEIN AI GENE TRANSCRIPTION.  
 CC BINDS TO DNA SITE A (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR2 SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X76653; CA54096.1; -  
 DR HSSP: P19793; 2NL.  
 DR MGD: MG1:1352452; NR2F2.  
 DR InterPro: IPR000536; -  
 DR InterPro: IPR001628; -  
 DR InterPro: IPR001723; -  
 DR InterPro: IPR003068; -  
 DR Pfam: PF00104; hormone\_rec; 1.

DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR PRINTS: PR00398; STRDHOMONER.  
 DR PRINTS: PR01282; COUPINFACOR.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Transcription regulation; DNA-binding; Nuclear protein;  
 KM Zinc-finger.  
 FT DOMAIN 71 75 POLY-GIN.  
 FT DNA\_BIND 79 144 NUCLEAR\_RECEPTOR-TYPE.  
 FT ZN\_FING 79 99 C4-TYPE.  
 FT ZN\_FING 115 139 C4-TYPE.  
 FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION.  
 SQ SEQUENCE 414 AA; 45597 MW; CB35C021B3127A99 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVPGSQ 6  
 |||||  
 DB 14 EVPGSQ 19

RESULT 16  
 COT2\_RAT STANDARD; PRT; 414 AA.  
 ID COT2\_RAT  
 AC 009018;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE COUP TRANSCRIPTION FACTOR 2 (COUP-TF2) (COUP-TF II) (APOLIPROTEIN AI  
 DE REGULATORY PROTEIN-1) (ARP-1) (OVALBUMIN UPSTREAM PROMOTER BETA  
 DE NUCLEAR RECEPTOR) (COUPB).  
 GN NR2F2 OR TFCOUP2 OR ARP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Plutitary;  
 RA Boutin J.M., Ronsin B., Devost D., Lipkin S.M., Rosenfeld M.G.,  
 RA Morel G.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REGULATION OF THE APOLIPROTEIN AI GENE TRANSCRIPTION.  
 CC BINDS TO DNA SITE A (BY SIMILARITY).  
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR2 SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF003944; AAB61297.1; -  
 DR InterPro: IPR000536; -  
 DR InterPro: IPR001628; -  
 DR InterPro: IPR001723; -  
 DR InterPro: IPR003068; -  
 DR Pfam: PF00104; hormone\_rec; 1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR PRINTS: PR00398; STRDHOMONER.  
 DR PRINTS: PR01282; COUPINFACOR.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Transcription regulation; DNA-binding; Nuclear protein;  
 KM Zinc-finger.

FT DOMAIN 71 75 POLY-GLN.  
 FT DNA\_BIND 79 144 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 79 99 C4-TYPE.  
 FT ZN\_FING 115 139 C4-TYPE.  
 FT DOMAIN 337 414 IMPORTANT FOR DIMERIZATION.  
 SQ SEQUENCE 414 AA; 45601 MW; 5AA05165DD3F675A CRC64;

Query Match 79.5%; Score 31; DB 1; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVPGSO 6  
 DB 14 EVPGSO 19

RESULT 17  
 AROC\_NEUCR STANDARD; PRT; 432 AA.  
 AC 012640;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 35, Last annotation update)  
 DE CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE).  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_Taxid=5141;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95386486; PubMed-7657620;  
 RA Henstrand J.M., Amrhein N., Schmid J.;  
 RT "Cloning and characterization of a heterologously expressed  
 RT bifunctional chorismate synthase/tryptophan synthase from Neurospora  
 RT crassa";  
 RL J. Biol. Chem. 270:20447-20452(1995).  
 CC -1- FUNCTION: BIFUNCTIONAL ENZYME THAT POSSESSES CHORISMATE SYNTHASE  
 CC AND INTRINSIC FLAVIN REDUCTASE ACTIVITY, IT USES NADPH TO REDUCE  
 CC FMN.  
 CC -1- CATALYTIC ACTIVITY: 5-O-(1-CARBOXYVINYL)-3-PHOSPHOSHIKIMATE -  
 CC CHORISMATE + ORTHOPHOSPHATE.  
 CC -1- COFACTOR: REDUCED FLAVIN.  
 CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
 CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
 CC -1- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: U25818; AAC49056.1; -  
 CC InterPro: IPR000453; -  
 DR Pfam: PF01264; Chorismate\_synth\_1.  
 DR PROSITE: PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
 DR PROSITE: PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
 DR PROSITE: PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
 KW Lyase; Aromatic amino acid biosynthesis; Oxidoreductase; NADP;  
 KW Multifunctional enzyme.  
 KW NP\_BIND 260 291 NADPH (POTENTIAL).  
 SQ SEQUENCE 432 AA; 45967 MW; 201A0B325C406F0C CRC64;

Query Match 79.5%; Score 31; DB 1; Length 432;  
 Best Local Similarity 85.7%; Pred. No. 41;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7

DB 274 EVPGSIH 280

RESULT 18  
 IL7R\_MOUSE STANDARD; PRT; 459 AA.  
 AC P16872;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE INTERLEUKIN-7 RECEPTOR ALPHA CHAIN PRECURSOR (IL-7R-ALPHA).  
 GN IL7R.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90199875; PubMed-2317865;  
 RA Goodwin R.G., Friend D., Ziegler S.F., Jerzy R., Falk B.A.,  
 RA Gimble S., Cosman D., Dower S.K., March C.J., Namen A.E., Park L.S.;  
 RT "Cloning of the human and murine interleukin-7 receptors:  
 RT demonstration of a soluble form and homology to a new receptor  
 RT superfamily";  
 RL Cell 60:941-951(1990).  
 CC -1- FUNCTION: RECEPTOR FOR INTERLEUKIN-7.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A COMMON GAMMA CHAIN.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: SPLEEN, THYMUS, AND FETAL LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: M29697; AAA39304.1; -  
 DR PIR: D34791; D34791.  
 DR MGD: MG1:96562; 117r.  
 DR InterPro: IPR000950; -  
 DR InterPro: IPR001777; -  
 DR InterPro: IPR002465; -  
 DR Pfam: PF00041; fn3; 1.  
 DR PROSITE: PS01355; HEMATOPO\_REC\_S\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 459 INTERLEUKIN-7 RECEPTOR ALPHA CHAIN.  
 FT DOMAIN 21 239 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 240 264 POTENTIAL.  
 FT DOMAIN 265 459 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 128 225 FIBRONECTIN TYPE-III.  
 FT DOMAIN 184 189 SER/THR-RICH.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD\_RES 282 282 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 SQ SEQUENCE 459 AA; 51704 MW; CC06A5CE95543849 CRC64;

Query Match 79.5%; Score 31; DB 1; Length 459;  
 Best Local Similarity 85.7%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7  
 DB 47 EVDGSOH 53

```

RESULT 19
ID BGAL_THEET STANDARD; PRT; 743 AA.
AC P77989;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN LACZ OR LACZ.
OS Thermoaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group.
OC Thermoaerobacter group; Thermoaerobacter.
OX NCBI_TaxID=1757;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33223 / 39E;
RA Zverlov V.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y08557; CAA69850.1; -.
CC DR InterPro: IPR001649; -.
CC DR Pfam: PF00703; Glyco_hydro.2; 1.
CC DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
CC DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
CC KM Hydrolase; Glycosidase.
CC FT ACT_SITE 388 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 453 NUCLEOPHILE (BY SIMILARITY).
CC SO ACT_SITE 453
CC FT ACT_SITE 453
CC SO SEQUENCE 743 AA; 85796 MW; FE01FE517E51DFC CRC64;

Query Match 79.58; Score 31; DB 1; Length 743;
Best Local Similarity 71.48; Pred. No. 72;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVPGSOH 7
DB 346 EIPGMOH 352

RESULT 20
ID KPTA_PSEAE STANDARD; PRT; 182 AA.
AC Q91776;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE RNA 2'-PHOSPHOTRANSFERASE (EC 2.7.7.6).
GN KPTA OR PA0054.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RL MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Boltman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;

```

```

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: REMOVES THE 2'-PHOSPHATE FROM RNA VIA AN INTERMEDIATE IN
CC WHICH THE PHOSPHATE IS ADP-RIBOSYLATED BY NAD FOLLOWED BY A
CC PRESUMED TRANSESTERIFICATION TO RELEASE THE RNA AND GENERATE ADP-
CC RIBOSE 1"-2"-CYCLIC PHOSPHATE (APPR>P). MAY FUNCTION AS AN ADP-
CC RIBOSYLASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE KPTA / TPT1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004445; AAC03444.1; -.
CC DR Transferrase; NAD.
CC KM Transferrase; NAD.
CC SO SEQUENCE 182 AA; 20024 MW; 2E3B121021A9F15E CRC64;

Query Match 76.98; Score 30; DB 1; Length 182;
Best Local Similarity 83.34; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VPQSOH 7
DB 118 VPQSRH 123

RESULT 21
ID RPOA_THEMA STANDARD; PRT; 336 AA.
AC Q9X112;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE ALPHA CHAIN) (RNA POLYMERASE ALPHA SUBUNIT).
GN RPOA OR TM1472.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RL MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson M.C., Ketchum K.A.,
RA McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton M.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
GN Nature 399:323-329(1999)
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN (BY SIMILARITY).
CC -1- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP, WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
CC EMBL: AE001798; AAD36540.1; -  
CC TIGR: TM1472; -  
CC InterPro: IPR001700; -  
CC Pfam: PF01000; RNA\_POL\_A\_bac; 1.  
CC Transferrase: Transcription; DNA-directed RNA polymerase.  
CC SEQUENCE 336 AA; 38600 MW; E361ED8D4408CF2 CRC64;  
SQ

Query Match 76.9%; Score 30; DB 1; Length 336;  
Best Local Similarity 71.4%; Pred. No. 52;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYPGSOH 7  
|||1:1  
Db 253 EYPASEH 259

RESULT 22  
AROC\_YEAST STANDARD; PRT; 376 AA.  
AC P28777;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE  
DE PHOSPHOLIMASE).  
GN ARO2 OR YGL148W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID:4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE-92114793; PubMed-1837329;  
RA Jones D.G.L., Reusser U., Braus G.H.;  
RT "Molecular cloning, characterization and analysis of the regulation  
RT of the ARO2 gene, encoding chorismate synthase, of Saccharomyces  
RT cerevisiae.";  
RL MOL. MICROBIOL. 5:2143-2152(1991).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1769;  
RX Voelt M., Deifoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
RT TTP1, MRF1 genes and six new open reading frames.";  
RL Yeast 13:177-182(1997).  
CC -1- CATALYTIC ACTIVITY: 5-O-(1-CARBOXYVINYLYL)-3-PHOSPHOSHIKIMATE -  
CC CHORISMATE + ORTHOPHOSPHATE.  
CC -1- COFACTOR: REDUCED FLAVIN.  
CC -1- PATHWAY: SEVENTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN  
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -1- INDUCTION: BY AMINO ACID STARVATION.  
CC -1- SIMILARITY: BELONGS TO THE CHORISMATE SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
CC EMBL: X60190; CAA42745.1; -  
CC

DR EMBL: X99960; CAA68214.1; -  
DR EMBL: Z72670; CAA96860.1; -  
DR PIR: S17246; S17246.  
DR SGD: S0003116; ARO2.  
DR InterPro: IPR000453; -  
DR Pfam: PF01264; Chorismate\_synth\_1.  
DR PROSITE: PS00787; CHORISMATE\_SYNTHASE\_1; 1.  
DR PROSITE: PS00788; CHORISMATE\_SYNTHASE\_2; 1.  
DR PROSITE: PS00789; CHORISMATE\_SYNTHASE\_3; 1.  
DR Lyase; Aromatic amino acid biosynthesis.  
SQ SEQUENCE 376 AA; 40838 MW; AF3AF65605B91E8E CRC64;  
SQ

Query Match 76.9%; Score 30; DB 1; Length 376;  
Best Local Similarity 83.3%; Pred. No. 58;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 YPGSKH 7  
||||1  
Db 266 YPGSKH 271

RESULT 23  
IL7R\_HUMAN STANDARD; PRT; 459 AA.  
ID IL7R\_HUMAN  
AC P16871;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE INTERLEUKIN-7 RECEPTOR ALPHA CHAIN PRECURSOR (IL-7R-ALPHA) (CDW127)  
DE (CD127 ANTIGEN).  
GN IL7R.  
OS Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-B-cell;  
RX MEDLINE-90199875; PubMed-2317865;  
RA Goodwin R.G., Friend D., Ziegler S.F., Jerzy R., Falk B.A.,  
RA Gimble S., Cosman D., Dower S.K., March C.J., Namen A.E., Park L.S.;  
RT "Cloning of the human and murine interleukin-7 receptors:  
RT demonstration of a soluble form and homology to a new receptor  
RT superfamily.";  
RL Cell 60:941-951(1990).  
CC -1- FUNCTION: RECEPTOR FOR INTERLEUKIN-7.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A COMMON GAMMA CHAIN.  
CC -1- SURCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN AND ALSO SECRETED.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; H20 (SHOWN HERE), H1 AND  
CC H6/SECRETED; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
CC -1- SIMILARITY: CONAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
CC -1- DATABASE: NAME-PROW: NOTE-CD guide CD127 entry:  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd127.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
CC EMBL: M29696; AAA59157.1; -  
CC PIR: A34791; A34791.  
DR PIR: B34791; B34791.  
DR PIR: C34791; C34791.  
DR MIM: 146661; -  
DR InterPro: IPR000950; -  
DR InterPro: IPR001777; -  
DR Pfam: PF00041; fn3; 1.  
DR





OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ST. LOCIA;  
 RX MEDLINE=99250256; PubMed=10231575;  
 RA Tataronov A., Saez A.G., Ayala F.J.;  
 RT "A compact gene cluster in *Drosophila*: the unrelated *Cs* gene is  
 RL compressed between duplicated *amd* and *Ddc*."; Gene 231:111-120(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pauley J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Hemington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."; Science 287:2185-2195(2000).  
 RL [3]  
 RP SEQUENCE OF 1-38 AND 102-346 FROM N.A.  
 RX MEDLINE=86038375; PubMed=3478553;  
 RA Eveleth D.D., Marsh J.L.;  
 RT "Overlapping transcription units in *Drosophila*: sequence and  
 RL structure of the *Cs* gene."; Mol. Gen. Genet. 209:290-298(1987).  
 RN [4]  
 RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RX MEDLINE=86284963; PubMed=2874495;  
 RA Spencer C.A., Gietz R.D., Hodgetts R.B.;  
 RT "Overlapping transcription units in the dopa decarboxylase region of  
 RL *Drosophila*."; Nature 322:279-281(1986).  
 RN [5]  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=86165362; PubMed=3007242;  
 RA Spencer C.A., Gietz R.D., Hodgetts R.B.;  
 RT "Analysis of the transcription unit adjacent to the 3'-end of the dopa

RT decarboxylase gene in *Drosophila melanogaster*."; RL Dev. Biol. 114:260-264(1986).  
 CC -1- FUNCTION: HAS A NONVITAL FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- TISSUE SPECIFICITY: LOW LEVELS SEEN IN ADULT HEADS, THORAX, ABDOMEN AND OVARIES. HIGH LEVELS IN TESTES.  
 CC -1- DEVELOPMENTAL STAGE: ABUNDANT IN EMBRYOS AND ADULTS.  
 CC -1- CAUTION: ORF1 AND ANON-37CS TRANSCRIPTS FROM REF.3 ARE NOW KNOWN TO BE FRAGMENTS OF THE COMPLETE ANON-37CS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF091328; AAC67581.1; -;  
 DR EMBL; AE003661; AAF53761.1; -;  
 DR EMBL; X05991; CAA29405.1; ALT\_SEQ.  
 DR EMBL; X05991; CAA29408.1; ALT\_SEQ.  
 DR PIR; S01105; S01105.  
 DR FLYBase; FBgn0002036; CG10561.  
 DR CONFLICT 2  
 FT CONFLICT 234 MISSING (IN REF. 2).  
 FT CONFLICT 234 R -> L (IN REF. 2).  
 FT CONFLICT 458 G -> D (IN REF. 1).  
 FT CONFLICT 475 G -> S (IN REF. 1).  
 FT SEQUENCE 504 AA; 56245 MW; ED7942B11F7A6080 CRC64;

Query Match 76.9%; Score 30; DB 1; Length 504;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 EYPSGSH 7  
 DB 372 QVPSOSH 378

Search completed: July 16, 2001, 16:44:03  
 Job time: 462 sec



**THIS PAGE BLANK (USPTO)**

From: Ford, Vanessa  
Sent: Tuesday, July 02, 2002 1:30 PM  
To: STIC-Biotech/ChemLib  
Subject: In re: 09786,648

Please search SEQ ID NOs: 2,3,4 and 5.  
Please run interference searches.

Vanessa L. Ford  
Biotechnology Patent Examiner  
Office: CM1 8D17  
Mailbox: CM1 8E12  
Phone: 703.308.4735

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

***The Pending database search results should not be left in the case because they contain data that is confidential.***

Searcher: P. Schneider  
Phone: 308-4292  
Location: CM1 6A03  
Date Picked Up: 7/3  
Date Completed: 7/8  
Searcher Prep/Review: 7  
Clerical: \_\_\_\_\_  
Online time: 6

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: 4  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where app  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: Compag  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**